Investigating Kissing to Duplex Dimer Transition Mechanism of HIV-1 SL1 by NMR

by

Xiaoyan Sun

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in The University of Michigan
2008

Doctoral Committee:

Associate Professor Hashim M. Al-Hashimi, Chair
Professor Carol A. Fierke
Professor Nils G. Walter
Assistant Professor Bruce A. Palfey

Acknowledgements

First of all, I would like to thank my advisor, Professor Hashim M. Al-Hashimi, for opening a door to an exciting scientific world for me, and I cannot thank him more for all his great advice, inspiration, encouragement, patience and support to keep me moving forward. I would also like to thank all the previous and present group members from the Al-Hashimi lab, for their kindness, friendship, and support. Especially, I would like to thank Dr. Qi Zhang, Dr. Melissa Getz, Anette Casiano, Max Bailor, Akash Bhattacharya and Gabrielle Todd for the great team work and collaboration in various projects and Alex Hansen for valuable scientific discussions and NMR tuition. I would like to thank Dr. Eric B. Bertelsen for his kindness and helps in teaching me molecular cloning techniques, Dr. Jingjie Mo for her collaboration and contributions on Mass Spectroscopy analysis in my work, and Dr. Alex Kurochkin for maintenance of NMR spectrometers and sharing the life experiences with me. I would like to thank my committee members, Professor Carol Fierke, Professor Nils Walter, Professor Erik Zuiderweg for their insightful advice for my research and especially Professor Bruce Palfey for his exceptional great ideas, help and supports throughout my kinetics project. I want to thank all my family members, Mom, Dad and give most special thanks to my husband for their love, encouragement and support for every aspect of my life.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

A – adenine
C – cytosine
CT – constant time
COSY – correlated spectroscopy
CPMG – Carr Purcell Meiboom Gill experiment
DNA – deoxyribonuclei acid
DD – duplex dimer of SL1
DIS – dimerization initiation site
E-SL1 – elongated stem loop 1
FID –free-induction decay
G – guanine
GDO (ϑ) –generalized degree of order
HIV-1 – human immunodeficiency virus type 1
HSQC –heteronuclear single quantum coherence
INEPT – insensitive nuclei enhanced by polarization transfer
IPAP – in phase, anti-phase detection scheme
J3/4 – junction region between helix 3 and 4
KD – kissing dimer of SL1
MES - 2-(N-Morpholino)ethanesulfonic acid
miRNA – micro RNA
mRNA – messenger RNA
NAIM –nucleotide analog interference mapping
NcRNA – non-coding RNA
NCp – nucleocapsid protein
NMR – nuclear magnetic resonance
NOE –nuclear overhauser effect
NOESY –nuclear overhauser effect spectroscopy
PBS – primer binding site
PDB – protein data bank
PEG – polyethylene glycol
RDCs – residual dipolar couplings
RMSD – root mean squared deviation
RNA – ribonucleic acid
SL1 – stem loop 1 of HIV-1
SL1m – mutated stem loop 1
TOCY – total correlated spectroscopy
TROSY – transferse relaxation optimized spectroscopy
tRNA –transfer RNA
U – uridine

Abstract

Investigating Kissing to Duplex Dimer Transition Mechanism of HIV-1 SL1 by NMR

by

Xiaoyan Sun

Chair: Hashim M. Al-Hashimi

As a common feature to retroviruses, human immunodeficiency virus (HIV) packages two identical copies of its single stranded RNA genome (gRNA) that hold together at its 5'-end. gRNA dimerization is initiated by stem loop 1 (SL1), which consists of a highly conserved asymmetric internal loop and a GC-rich self-interacting palindromic apical loop that can drive dimerization by forming a meta-stable kissing dimer. During maturation, the kissing dimer undergoes a transition catalyzed by the viral nucleocapsid protein (NCp) into a thermodynamically more stable duplex. Both dimerization and structural isomerization between the kissing and duplex dimer are critical for viral replication and packaging. While SL1 and NCp have been the focus of many studies, the mechanism of the NCp dependent SL1 dimerization and isomerization remains poorly understood.

This dissertation describes the characterization of SL1 structural dynamics, its Mg^{2+} and NCp binding properties, and the time-course of the kissing to duplex transition using high resolution nuclear magnetic resonance (NMR) spectroscopy. Initial studies were conducted on the conformational properties of the internal loop of SL1. Subsequently, we characterized the corresponding properties in kissing and duplex SL1 dimers along with their interaction with NCp and followed site-specifically the time-course of the kissing to duplex transition using time-resolved NMR.

We observe three types of motions that may promote conversion of the kissing dimer into its duplex form: (i) diffusion-limited nanosecond collective helix motions about the G-AGG internal loop that may help bring strands from distinct monomeric units into a proper register; (ii) a secondary structural switch occurring at μs-ms timescales which leads to partial melting of the upper-helix; and (iii) looping-in-and-out hinge motions of adenine residues in the apical loop that may help bring strands from different monomers into close spatial proximity. All three classes of motions are significantly dampened by Mg^{2+} which likely serves to make the transition strongly dependent on NCp. The NCp interacts with the internal loop and the apical loop of kissing dimer. Our results suggest that NCp stabilizes an alternative SL1 conformation, likely involving a quadruplex geometry, prior to transitioning in a single step into the duplex dimer.

Chapter 1

Introduction

The human immunodeficiency virus (HIV) is the causative agent of the acquired immunodeficiency syndrome (AIDS), one of the deadliest epidemic in the history of mankind, which was first recognized and identified in the early 1980s (1). HIV-1 and HIV-2 are the only two species of HIV that can infect humans: HIV-1 transmits more easily and is more hazardous than HIV-2, which is the cause of the majority of HIV infections globally (2,3). The fight against HIV has mainly focused on targeting essential viral protein enzymes, such as reverse transcriptase, protease, and integrase during different stages of the HIV lifecycle (4). Many drugs on the market have been used to target these viral proteins, such as Zidovudine, which is a nucleoside analog and reverse transcriptase inhibitor (NARTIs or NRTIs) (5); Saquinavir (Hoffmann–La Roche™), which inhibits protease; and aportmanteau inhibitor, which is a combination of an HIV reverse transcriptase inhibitor and an integrase inhibitor (6). There is, however, an imperative need for developing and investigating new anti-viral drug targets due to the high mutation and recombination rates of HIV-1 and its ability to self-select drug-resistant strains through high replication levels (7). Highly conserved, structured RNA elements in the 5’ non-coding region of the retroviral genomic RNA, which regulate key steps in the HIV-1 life cycle, have been proposed as potential drug targets (8,9).

In this thesis, we structurally, dynamically, and functionally characterize one such promising candidate for anti-HIV therapies, the stem loop l (SL1) RNA domain (10) and its interaction with the viral nucleocapsid protein (NC), which is derived from the NC domain of the precursor Gag polyprotein by proteolytic processing, binds residues in the apical loop of SL1 (11-13), which play essential roles in HIV-1 RNA genome dimerization.

1.1 RNA genome dimerization in the life cycle of retroviruses

1.1.1 Dimerization in retroviruses

Retroviruses carry their genetic material in the form of RNA rather than DNA and use the enzyme reverse transcriptase to reverse transcribe the RNA into DNA. With the exception of spumaretroviruses, all retroviruses have two nearly identical single stranded genomic RNA (gRNA) molecules, which range between 8,000 to 11,000 nucleotides in size and contain a 5’ cap and 3’ poly(A) sequence as eukaryotic mRNAs (14). The two homologous RNAs are coated by 2,000 to 3,000 molecules of NC protein, an RNA chaperone (15), and are linked near the 5’ end (16) via non-covalent interactions to form a dimeric structure. This feature of the retroviral RNA genome was first observed by sedimentation analysis in a variety of retroviruses, such as Rous sarcoma virus (RSV), avian myeloblastosis virus (AMV), and murine leukemia virus (MLV) in the 1960s

(17,18). It was subsequently visualized by electron microscopy (EM) in a variety of viruses, including HIV (19-22).

Studies showed that during viral maturation, the HIV-1 gRNA dimer undergoes a transition into a more heat-stable form (23-29). This transition was coupled to maturation of the HIV-1 particle which features cleavage and release of the NC protein from the polyprotein Gag (30). Mutational studies showed that the release of NC protein from Gag is critical for the maturation of the RNA dimer (31); in addition, viral dimeric RNA enhances the proteolytic processing of retroviral polyproteins and stabilizes the viral particle architecture during virion maturation (32) (Figure 1-1). These results establish an as yet poorly understood correlation between RNA dimerization and maturation and virion core formation (33-36).

1.1.2 Dimeric linkage sequence in the HIV-1 genome

In HIV-1 the site of dimerization, termed the dimer linkage sequence (DLS), was found to be close to the 5' end of the RNA genome (20-22). Stem-loop 1 (SL1) was shown to be a key component of the DLS structure. SL1 contains a self-complementary apical loop that can drive formation of a loose kissing dimer (24,37-39) (Figure 1-1). The loose kissing dimer can subsequently undergo a transition to a duplex conformation in which the two gRNA copies are held together by stronger inter-strand interactions (40) (Figure 1-1). This transition was shown to be stimulated by the NC protein (41-44).

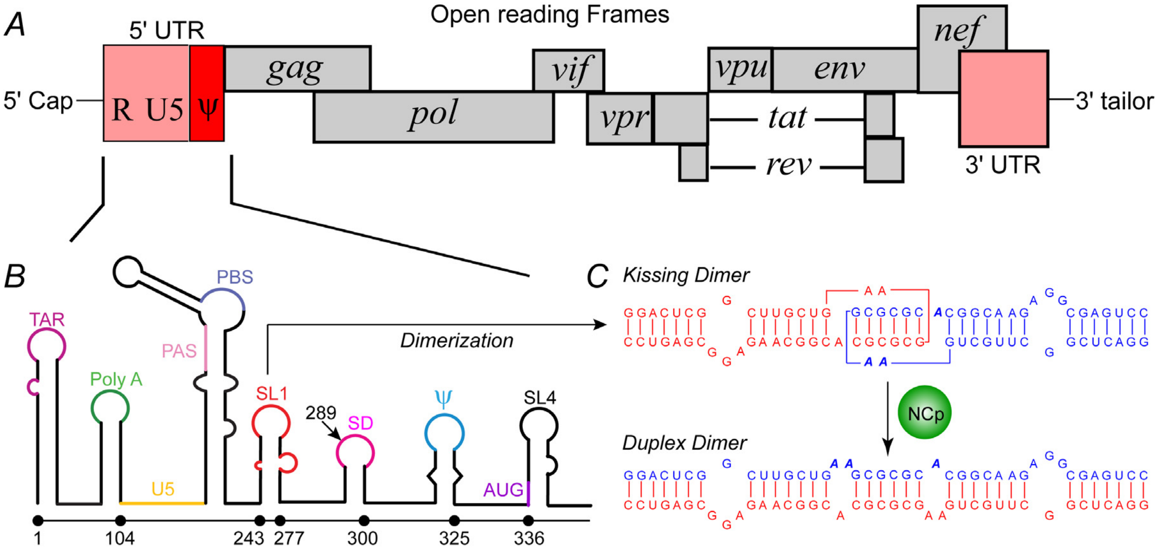

Figure 1-1. SL1 RNA from HIV-1 genome. Shown are (A) organization of HIV-1 genome, (B) the secondary structure model of the HIV-1 untranslated leader RNA, and (C) the conformational changes essential to SL1 function.

1.1.3 Role of dimerization in HIV replication

Dimerization is believed to play diverse roles in viral replication. Early *in vivo* data showed that only HIV-1 viruses with dimeric gRNAs are capable of efficient infectivity (45,46). Dimerization serves as a “hotspot” for retroviral genome recombination which is critical to the high variation ability of HIV (24,47-51) (Figure 1-1). For example, templates containing the dimerization site had 4-fold higher strand transfer efficiency than the non-dimeric control during reverse transcription (47,48). Heat dissociation of the dimeric viral genome strongly suppressed the reverse transcription (50,51). In addition, DLS acts as a regulatory switch for permitting or restricting the packaging of the gRNA (52-57). For example, partial deletion or mutations of DLS leads to a significant reduction of gRNA packaging and the efficiency of encapsidation (24,39,58,59). Moreover, the dimeric gRNA is believed to be a scaffold for multimerization of the precursor Gag protein, which is a crucial requirement for the formation of virion particles during virion assembly (32,60).

1.1.4 Other HIV-1 conformational transitions linked to SL1 and genome dimerization

The SL1 mediated dimerization is believed to be linked and possibly further regulated by another structural transition involving larger fragments of the 5' UTR (Figure 1-2). The 5' non-coding RNA (~290 nt) has been shown to adopt two distinct conformations that migrate at different rates on native polyacrylamide gels (61,62). Secondary structure prediction and chemical probing suggest that the faster migrating species adopts an alternative, extended “long-distance interaction” (LDI) conformation that cannot be converted into dimers because the GC-rich SL1 loop is masked by base-pairing with the poly(A) hairpin (Figure 1-2) (62). The slowly migrating species is believed to be the conventional “branched multiple hairpin” (BMH) conformer that is able to dimerize since the SL1 GC-rich loop is exposed (Figure 1-2) (62). The LDI can be converted into BMH by addition of NC protein or Mg^{2+} (61,62). This conformational switch is believed to further regulate dimerization of the HIV-1 genome (61,63).

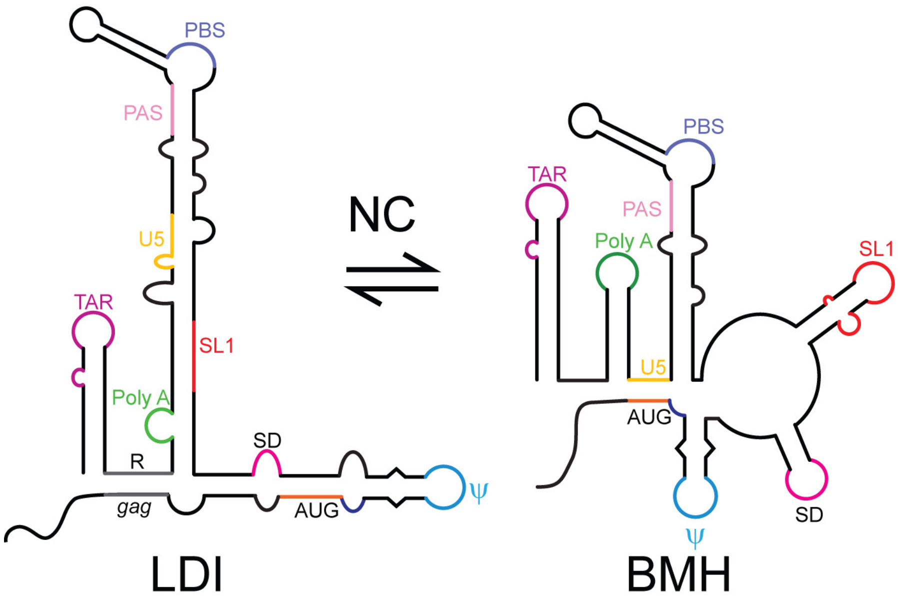

Figure 1-2. Two alternative structures of HIV genome leader RNA.

1.2 Stem-loop 1: A highly conserved RNA element that plays essential roles in the dimerization of the HIV-1 genome

1.2.1 Variation of SL1 RNA in HIV-1

SL1 is strongly conserved among different strains of HIV-1, which can be classified into three groups: the "major" group (M), the "outlier" group (O), and the "new" group (N). Phylogenetic analyses of the nucleotide sequences of a large number of HIV-1 isolates have identified the existence of at least eight different genetic subtypes of HIV-1 group M, which have been designated A through K, corresponding to different geographical distributions (64-67). Subtype A is mainly found in North-West Africa and South-East Asia (Figure 1-3A). Subtypes B and D are more closely related compared to other stains, and are widespread over the “Western” world, including North and Central America, Europe and Australia (Figure 1-3A). Subtype F and G were initially identified in Central Europe and South Africa, and subtype H was initially identified in Thailand and is also present in the Central African Republic (Figure 1-3A). Subtype C is prevalent in India and recently was identified in southern Brazil. These genetic subtypes are approximately similar to each other and more than 90% of HIV-1 infections belong to this group M. The Group O viruses are distantly related to group M viruses, and have been recently identified in patients from Cameroon. Group O is comprised of a number of distinct genetic subtypes, which remain to be identified and characterized (68). The group N virus was discovered in 1998 in Cameroon and is extremely rare.

Most virulent HIV-1 SL1 isolates fall into two classes: subtype A and subtype B/D (UNAIDS, 2002†). Those SL1 RNA elements from the major HIV-1 group have a nine-nucleotide apical loop and a 100% conserved G-AGG internal loop (Figure 1-3B). The greatest variations are observed in the apical loop sequence as shown in Figure 1-3A. In addition to the 6 nt GC rich palindromic loop sequence, the most virulent subtypes (conservatively B, D) include three adenines in the apical loop, two of them located at the 5' end and one at the 3' end. In this thesis, we will focus on the most common and virulent HIV-1 subtype in North America: subtype B.

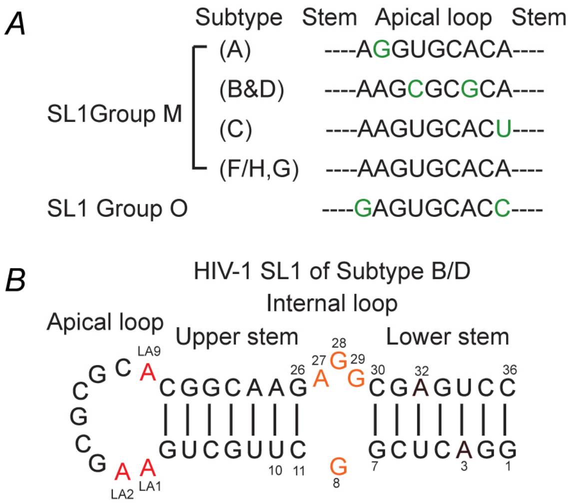

Figure 1-3. Secondary structure and variation of subtypes of HIV-1 SL1 RNA.

1.2.2 Key sequence elements of SL1 subtype B

SL1 contains two essential elements: an apical loop and an asymmetric internal loop. The apical loop contains a 6 nt self-complementary palindromic GC sequence which promotes spontaneous genome dimerization through formation of a loose kissing dimer that is held together by intermolecular loop-loop base-pairing, (41,69-71). As discussed above, only two out of 64 possible self-complementary sequences have been found in the SL1 apical loop among a variety of HIV isolates, raising the possiblity that a self-complementarity sequence is not the only essential feature of the apical loop structure. A variety of studies suggest that NC protein may bind the apical loop in the kissing dimerization. Additionally, the three unpaired adenines flanking the apical loop are important for dimerization. Deletion of these residues or their replacement with uridines eliminates the dimerization and strongly affects the RNA association rate (72). Moreover, the two adenines in the first and last position of the apical loop are strictly required for the formation of a stable loop-loop interaction (73).

The highly conserved asymmetric internal loop is also required for genome dimerization (74). Deletion of the internal loop and the lower stem of SL1 decreases the amount of the dimer *in vivo* and of proviral DNA synthesis by 85% (51). The internal loop also plays an essential and possibly dynamic role in the transition (or isomerization) between kissing and duplex dimers. Deletion of the internal loop only prevents formation of the duplex dimer (40). SL1 kissing dimers containing the internal loop can spontaneously convert into duplex dimers in the absence of NC protein when incubated at 55°C, whereas corresponding constructs lacking the bulge cannot form duplex dimers even in the presence of NC protein (26,27,40). The internal loop may also serve as an important NC binding site. Chemical probing experiments show that guanine residues G28 and G29 (Figure 1-3B) are accessible in the context of the 206 nt Ψ HIV-1 RNA (75). Footprinting data on a 401 nt fragment of the HIV-1 RNA leader shows that G28 and G29 are both strongly accessible in the free RNA and strongly protected upon binding to GST-tagged Gag or NC (76). Fluorescence experiments show that NC protein binds not only to the SL1 apical loop but also to the internal loop (11,77). More recent studies using electrospray ionization-Fourier transform mass spectrometry show that the interaction between NC protein and the internal loop is key for inducing the kissing-duplex structural transition, whereas the interaction between NC protein and the apical loop can inhibit dimer formation (78).

1.2.3 Structure and metal binding properties of SL1 kissing dimer

Three-dimensional models for the SL1 subtype A and B in kissing dimer form were first proposed by Paillart and co-workers based on biochemical probing data (79). They reported the existence of a non-canonical base pair between the first 5’ end and the 3’ end adenines in both subtypes of kissing dimer. Moreover, they revealed a base triple in the subtype A kissing dimer that involve loop G2 base pairing with kissing loop base-pair C6-G5’, but this base triple was not found in the subtype B kissing dimer where instead loop A2 stacked below loop A2’(79). Subsequently, a ^{1}H NMR study reported a kissing dimer structure for a 19mer SL1 fragment from subtype A, which contained a wild type apical loop but a mutated stem (80). This structure revealed that 5’ unpaired adenines from the apical loop are stacked inside the helix adjacent to the G15-C20 base pair from the upper stem and form a non-canonical A1-A9 base pair consistent with previous studies (79). However, broadening of the loop G2 exchangeable proton was observed (80), which stood in contrast to the base triple proposed by Paillart and co-workers based on biochemical data (79). This was the beginning of what became a long string of discrepancies regarding the conformation of SL1 as summarized in Table 1-1.

The first high resolution structure for the SL1 subtype B kissing dimer was solved by NMR (81). The 23mer kissing dimer featured a wild type apical loop and upper stem (81). The SL1 kissing dimer structure was intrinsically bent exhibiting limited hinge-like flexibility at the junction between the rigid helical segments. In this structure, the kissing loops are oriented roughly perpendicular to the axis of the two helical stems and the unpaired 5' and 3' end loop adenines (LA) adopt looped in conformations. In addition, the structure revealed a new intermolecular base triple stacking interaction which appeared critical for stabilizing the dimer. In this base triple, LA2 crosses the interface to stack with 5' LA1' on the opposite strand while disrupting the G15'-C20' base pair adjacent to the apical loop by stacking onto C20'. This feature differed from that reported by Paillart et. al. The 3' LA9 is looped in and stacked against LC8 or occasionally with C20 (81). These studies were conducted in 10 mM sodium phosphate buffer and 100 mM NaCl, pH 7.2 at 15-25°C (81).

Soon thereafter, crystal structures of the 23mer SL1 kissing dimer of subtype A and B were reported under high ionic strength conditions (5 mM $MgCl_2$, 150 mM KCl, and 20 mM Na cacodylate, pH 7.0). Interestingly, the two subtypes folded into identical conformations, in which the apical loop base-pairs form a helix that is in perfect coaxial alignment with the helical stems. Surprisingly, the 5' LA1 and LA2 (in subtype A) / LG2 (in subtype B) were stacked with each other in a bulged-out conformation, and the 3'

LA9 stacked inside the helix (82). Subsequent X-ray structures of the SL1 subtype F kissing dimer also revealed that LA9 is perfectly stacked within the helix, whereas the LA1 and LA2 are stacked extra-helically (83).

Clearly, there are discrepancies between the above NMR and X-ray structures of the SL1 kissing dimer. While the NMR structure (81) shows a kissing loop roughly perpendicular to the stem axis and three adenines stacked inside the helix, the crystal structures shows a perfect coaxial alignment for the helices with the 5' purines stacked on each other in a bulged-out conformation and the 3' purine stacked within the helices as observed in the NMR structure (82,83). Subsequently, a new NMR structure of SL1 subtype B kissing dimer was solved based on a large number of NOEs. The structure showed coaxial alignment for the helices in agreement with the X-ray structure and then significantly differed from the previous NMR structure (81). However, it revealed a bulged-in conformation for the two 5' unpaired adenines similar to the other NMR structure but contradicting the X-ray structure (84). Another NMR structure of SL1 subtype A kissing dimer, derived using a divide-and-conquer approach (85), reported an A-form like linear structure with all unpaired purines in a bulged-in conformation. Several studies suggest that the two 5' unpaired As likely exist in a dynamical equilibrium between bulged-in and bulged-out states. Additionally, a Mg^{2+} dependent bulging-out mechanism has been proposed (86).

Table 1-1. Summary of the structure of unpaired adenines in the kissing dimer loop

X-ray Kissing 23mer	Subtype A (MAL)			Subtype B (LAI)			Subtype F		
Unpaired loop res.	A1	G2	A9	A1	A2	A9	A1	A2	A9
In-out state	out	out	in	out	out	in	out	out	in
Structure features	A1/G2 stack in pairs, separated by one base from other strand (open)			A1 and A2 stacking in pairs (open)			A1, A2 and A1', A2' continuously stack (close)		

NMR Kissing 23mer	Subtype B (Mujeeb)			Subtype B (Keiken)			Subtype A (Dardel)			Subtype F		
Unpaired loop res.	A1	A2	A9	A1	A2	A9	A1	G2	A9	A1	A2	A9
In-out state	in	in	in	in	in	in	in	in	in	in	in	in
Structure features	A2' stack between A1 and C20.G15-C20 unpaired			A2' stacks on A1 A1' stacks on A2 G15-C20 paired			A1/G2 stack inside helix (no structure)			A1 stacks on G15; A2 stacks on A2' (flexible?)		

1.2.4 Structure and metal binding properties of duplex dimer

The first SL1 subtype A duplex structure (86) was solved by X-ray crystallography. In contrast to the subtype A kissing dimer (82), the duplex structure showed a 'base-grip' motif formed by two extra-helically bulged adenines on each strand, flanked by two G2–A9' mismatches. Three magnesium ions were observed around the bulged-out As, which are directly involved in forming and stabilizing the duplex structure. In the same year, the first NMR structure of a 23mer SL1 (subtype B) duplex dimer (87) was solved. It showed a unique conformation with a zipper-like motif involving LA1, LA2 and LA9' from other strand partially stacking within the helix. Yet another NMR structure of SL1 (subtype B) duplex dimer reported unpaired loop adenines stacking in the helix with a collinear base stacking structure (88).

The only structure of the SL1 subtype B duplex dimer that includes the internal loop and lower stem reported an overall A-form structure with all three flanking adenines situated inside the helix (89). The structure of the SL1 subtype A duplex dimer was solved by a divide-and-conquer approach (85). The overall shape of the two duplex dimer structures described above are distinctly different: the subtype A duplex dimer has a single bend in the center of the palindromic apical loop, while the subtype B has two bends associated with the internal loop of SL1 without the bend in the middle of the palindromic loop.

These conflicting results raise the possibility that intrinsic flexibility of SL1 loop residues plays an essential role in its functions. Many molecular dynamics (MD) simulation studies on kissing and duplex dimers have provided insights into the dynamics of SL1 kissing and duplex dimers. Many of these studies confirm the existence of differential mobility at different parts of the SL1 which again differ across different sub-types and in the kissing and duplex dimer, again with unique metal binding properties. The looping in and out of apical loop As has also been observed in MD studies that started with a looped in conformation derived from an X-ray structure (90-94).

Table 1-2. Summary of the structure of unpaired adenines in the duplex dimer loop

X-ray Duplex 23mer	Subtype A (MAL)			Subtype B (LAI)		
Unpaired loop res.	A1	G2	A9	A1	A2	A9
In-out state	out	in	in	out	out	in
Structure features	G2 and A9 base paired and stack within helix, A1 is bulge out			A1 and A2 stacking out		

NMR Duplex	Subtype B (one with internal bulge) (both from James TL)			Subtype B (1JU1) 23mer			Subtype F (Baba) 23mer		
Unpaired loop res.	A1	A2	A9	A1	A2	A9	A1	A2	A9
In-out state	In	in	in	in	in	in	in	in	in
Structure features	A9’ stacks between A1 and A2, A1 stacked slightly against G15			A1, A9’ and A2 partially stacked, form zipper-like motif, NOE of A9’ connect with C8 and C20			A9’- A1base paired and A2 stacks on A1		

1.2.5 The asymmetric internal loop

The internal loop has been shown to be crucial for the kissing to duplex transition (40,95-97) induced either by NC protein at 37°C or even spontaneously by increasing the temperature to 55°C. However, structural information on the internal loop in the intact kissing dimer is still lacking. Existing structural studies on the internal loop have used SL1 constructs with modified apical loops to prevent dimerization. In these constructs, the 9 nt apical loop was replaced by non-self-complementary tetraloops, which are known to stabilize stem-loop structures.

Interestingly, differences in the conformational behavior have also been reported for the internal loop. On one hand, Greatorex et al. had to modify the wild-type G-AGG internal loop to a G-GGA loop in order to quench debilitating exchange broadening that is indicative of slow μs-ms motions (95). On the other hand, Lawrence et al. reported a stable structure for the internal loop, in which a guanine and an adenine from opposite strands form a G-A mismatch stacked on the two remaining guanines (G28, G29) which are staked within the helix (98). In another study, Yuan et al. reported that the internal loop exchanges between three conformations (99). Unlike the highly flexible internal loop reported by Greatorex et al, Baba et al. also believed a rather well structured molecule with some dynamics in the internal loop (85)

MD simulation studies were used to investigate the dynamics of the internal loop (97,100). Without an intact structure of the kissing dimer, they combined the X-ray structure of the 23mer SL1 kissing complex (82) and the internal loop structure of the SL1 mutant solved by NMR (98). Their results revealed that the internal loop is very flexible and destabilizes the upper stem (100). A new MD simulation recently reported that the internal loop displays conformational variability that are coupled with the flexibility of the 9 nt apical loop and causes large conformational fluctuations in the 7 bp stem linking the apical and internal loops (97).

Currently only one internal loop structure in the duplex conformation has been revealed by NMR (89). The G-AGG internal loop displays high conformational variability and exhibits a similar GA mismatch to that observed in the SL1 monomer (98). The two 3' guanines are exposed, one in the minor groove and another in the major groove side of the helix. While this is consistent with biochemical probing studies for SL1, it disagrees with Lawrence et al.'s report that these two G's stacked between the G–A mismatch and the lower stem which are not accessible for NC protein.

1.3 The Nucleocapsid protein

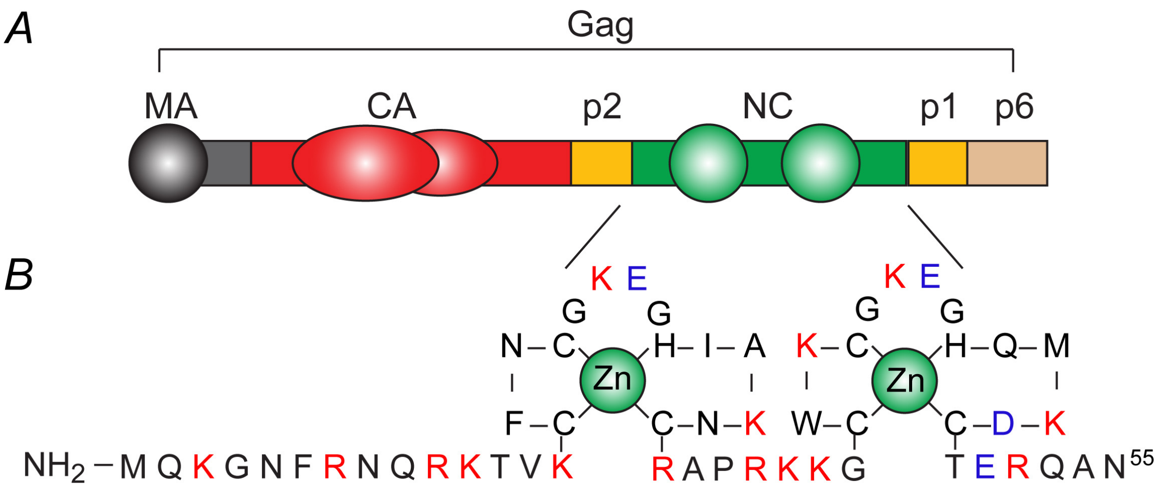

Figure 1-4. The origin and secondary structure of NC protein.

The NC protein is a key component in retroviral particles and takes on various functional roles in the retrovirus life cycle, such as the specific binding and annealing activities to viral RNA (101,102). The NC domain is the third element located at the C-terminus of the Gag polyprotein which is necessary and sufficient for assembly of retrovirus particles in eukaryotic cells (15,103). In retroviruses, the primary function of the NC domain is to bind specifically to the packaging signals and deliver a full length and most likely dimeric viral RNA genome into the assembling virion. Shortly after the virion is released from the host cell, the Gag polyprotein is cleaved into a series of functional small proteins, including the NC protein. This process is believed to be coupled with the formation of matured dimeric gRNA (104) (Figure 1-4A).

Retroviral NC protein is 55 amino acids in length and is highly basic with an unusual pI of 10.0 to 11.0. It nonspecifically binds to single stranded nucleic acids 5 to 10 nucleotides in length, which results in protection of viral genomes from nuclease attack and presumably compacts them within the virion core (105,106). Interestingly, dose-response studies consistently showed that a threshold concentration of the protein is required for demonstrable nucleic-acid-chaperone effects. For instance, the efficiency of reverse transcription increased 10-fold when the ratio of nucleotides per NC protein reaches 0.9 from 7 (107-109). It appears that there is little chaperone effect unless the nucleic acids are almost completely covered by the protein, which in turn raises the possibility that there is no turnover in the catalysis of conformational changes by NC protein. The need for saturating amounts of NC protein might also indicate that interactions between NC protein bound to nucleic acid play a role in the chaperone effect (110).

NC protein binds both to ss/ds DNA (111) and to ssRNA. NC protein presents an unusual nucleic acid chaperone activity and catalyzes the refolding of nucleic acids into conformations that have the maximal number of base pairs by lowering the energy barrier for breakage and reformation of base pairs. NC protein is believed to interact with nucleic acids so to induce disruption of base-pairs allowing refolding into alternative combinations. NC protein catalyzes a number of functionally important transitions, including the annealing of tRNA to primer binding sites of HIV-1, gRNA secondary structural transitions, DNA strand exchange during reverse transcription (112)(107)(113), and viral genome integration (114). Moreover, NC protein facilitates the initiation and elongation steps of proviral DNA synthesis and p55gag multimerization and incorportation of p160gag-pol into viruses. The affinity of the NC protein to RNAs follows the general order: retroviral RNA > mRNA > rRNA > PolyA. For retroviral RNA, the association constant is estimated to be ~$5x10^7$ M^{-1} (115,116) and ~$3x10^7$ M^{-1}

for HIV-1 RNA (117). In addition, competition binding between the cations and NC protein to RNA has been observed which indicated an electrostatic contribution to the binding process (117).

NC protein contains 55 amino acids and can be divided into an N-terminus region (residues 1-13, a basic region), a proximal zinc finger (residues 14-30), linker (residues 31-34, a strongly basic region), distal zinc finger (residues 35-51), and C-terminus (residues 52-55, prior to proteolytic cleavge) (118,119) (Figure 1-4B). The two zinc fingers belong to the CCHC-type family of structures (Cys-X_2-Cys-X_4-His-X_4-Cys, where X is a variable amino acid), which form very tight and rigid loops. The conserved hydrophobic and aromatic residues in the zinc finger form a hydrophobic cleft on the surface. Mutation of these residues leads to altered RNA packaging specificity (104,120-123) .

NC protein is required for efficient or complete formation of thermolabile viral RNA dimers and for refolding of thermolabile dimer into mature and thermostable dimers (124). It has been shown to catalyze the conversion between the kinetically trapped kissing and thermodynamically more stable duplex SL1 dimers *in vitro*. Additionally, recent studies provide evidence that gRNA dimerization can be inhibited by mutations in conserved glycines of either zinc fingers, and the N-terminus or the linker region (125). Interestingly, early filter-binding studies indicated that NC binds to intact SL1 with similar affinity as SL3 (200 nM) (75). More recent studies indicate that its affinity for SL1 monomer is actually 70% (for apical loop) – 80% (for internal loop) less than its affinity for SL3, which is 28 nM (11).

1.4 Proposed mechanisms for stem loop I kissing to duplex isomerization

The kissing to duplex transition is thought to involve a two-step transition: (i) formation of a thermodynamically metastable kissing dimer via a self-complementary GC-rich palindrome in the SL1 loop and (ii) isomerization of the loop-loop kissing dimer into the thermodynamically more stable duplex dimer (Figure.1-1C) (27,40). The transition can occur in an NC dependent manner or even spontaneously in the absence of NC upon incubation at 55°C. The mechanism of the kissing to duplex transition is still poorly understood. Four general types of mechanisms have been proposed: (i) fusion and re-association of stem helices (38,126) (Figure 1-5A), (ii) loop-loop helix extension (92) (Figure 1-5B), (iii) a ribozyme-like mechanism by cleavage and cross-religation (82) (Figure 1-5C) and (iv) a conversion mechanism that features a quadruplex intermediate (Figure 1-5D).

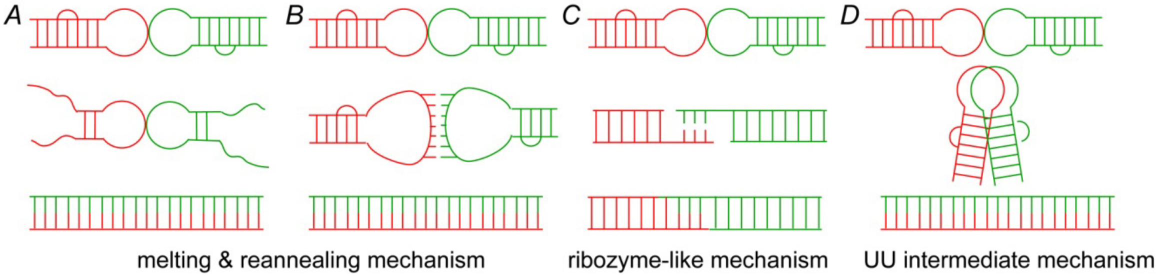

Figure 1-5. Proposed kissing to duplex conversion mechanisms.

Most evidence supports that the loop-loop interaction remains intact during the transition (73,77,92,126-128). The unpaired purines from the apical loop are believed to play essential roles in the transition. For example, it has been proposed that one of the unpaired purines is the first residue to reach the other side of the kissing complex and induces formation of new intermolecular base-pairing during the loop-loop extension mechanism (92). It has also been proposed that the unpaired purines are a hot spot for Mg^{2+} binding and cleavage in the ribozyme-like mechanism (82). It has been reported that one of the adenines can be protonated by NC protein in the kissing dimer, which goes on to form two or more dynamic ensemble states of the kissing dimer, allowing NC binding and activation of the kissing to duplex transition. In contrast, a recent electrospray ionization-Fourier transform ion cyclotron resonance mass spectroscopy study proposed that the NC protein that drives the transition does not bind the apical loop but rather the internal loop (12). The above discussion highlights how, despite numerous studies, the mechanism of SL1 isomerization remains poorly understood and the role, as well as binding sites, for NC protein as still elusive.

Table 1-3. Conversion Rate of kissing to duplex dimer in the presence of NC protein (NCp)

100 nM DIS24(GA)-DIS24(UC) kissing complex (77)			
75 nM NCp	$k_{conv} = 0.084 \pm 0.008\ min^{-1}$	100 nM NCp	$k_{conv} = 0.077 \pm 0.007\ min^{-1}$
150 nM NCp	$k_{conv} = 0.06 \pm 0.006\ min^{-1}$	250 nM NCp	$k_{conv} = 0.460 \pm 0.024\ min^{-1}$
400 nM NCp	$k_{conv} = 1.39 \pm 0.06\ min^{-1}$	500nM NCp	$k_{conv} = 2.2 \pm 0.15\ min^{-1}$

1.5 NMR methods for characterizing nucleic acids

1.5.1 A brief survey of nucleic acid NMR

NMR spectroscopy is one of two techniques that can be used to determine the three dimensional structure of macromolecules at atomic resolution (129). Compared with X-ray crystallography, NMR stands out due to its broad capabilities for investigating molecular dynamics, reaction kinetics, folding, and intermolecular interactions (130-134). Since the first applications of pulsed Fourier transform NMR in studies of tRNA in the 1970s, there has been an explosive growth in NMR studies of structure and dynamics of nucleic acids in solution and their interactions with metals, small molecules, proteins, and other nucleic acids. However, the structural homogeneity imposed by common A-form helical RNA elements, and the chemical similarity between only four nucleotides, leads to limited chemical shift dispersion, thus making it very difficult to assign resonances in congested spectra, particularly for ribose sugar protons. Continuous developments of multidimensional NMR methodology starting from 1970s, together with the technique development for preparing isotropically ^{13}C and ^{15}N labeled RNA samples at moderately high concentration [0.2-1 mM, 6-30 mg/ml], have paved the way for quantitative NMR studies of nucleic acids. To date, 814 of the total 1899 three-dimensional structures of nucleic acids reported in the protein data bank have been solved by NMR spectroscopy.

In addition to being able to provide more physiological relevant solution conditions for biomolecular structure determination, avoiding crystal packing forces that might have significant effects on flexible nucleic acids conformations, NMR offers many other advantages for characterizing biomolecular properties that are essential for their biological functions, such as the ability to quantitatively characterize internal motions over a wide range timescales from picosecond to seconds (135-138). For example, spin relaxation methods can be used to obtain information about internal motions occurring at pico-to-nanosecond timescales, whereas relaxation dispersion experiments can be used to access internal motions occurring at micro-to-millisecond timescales and can also be used to characterize “invisible” excited states of biomolecules (139-141). Residual dipolar couplings (RDCs) can provide long range angular information, characterize internal and domain-domain motions, and aid in generating motional trajectories of biomolecules occurring faster than millisecond timescales (142-147). Recently developed time-resolved NMR techniques can also be used to monitor biomolecular folding/unfolding process on the seconds timescale (148). The ability to provide information of both structure and dynamics at atomic resolution enables NMR a unique and powerful technique to characterize and understand the functions of biomolecules from comprehensive structural biology point of view.

1.5.2 Resonance assignments and structure determination

The chemical shift of a given resonance contains valuable information about its local chemical environment, and its assignment, i.e. identifying the belonging of a given resonance, is usually a requirement for further NMR study. A basic NOESY experiment on an unlabeled RNA molecule correlates all protons within a distance of 5 Å via through space dipolar-dipolar interaction. Due to the spatial proximity to each other, strong NOE cross peaks can be obtained between the H2 proton of adenine and the H3 imino proton of uracil in an A-U base pair and between the amino protons of cytosine and the H1 imino proton of guanine in a G-C base pair (149). Additionally, by applying long mixing time, cross peaks between the imino protons and H5 and H6 protons of cytosine and uracil also can be detected due to spin diffusion. Furthermore, detecting exchangeable (N-H) and nonexchangeable (C-H) NOE cross peaks using 2D/3D NOESY (^{15}N or ^{13}C filtered) experiments on isotopically labeled RNA samples provides unique and essential information to sequentially assign the proton resonances of the base pairs in RNA helix. Moreover, NOE cross peaks can also be used to establish both intra- and inter-residue correlations between base and sugar protons. In A-form helices, for instance, the base H8 in purines can be correlated to its own sugar H1' and the proceeding H1' in the 5' direction; therefore, an “NOE walk” can be used to assign H2, H6, H8, and H1' protons from the 3' to the 5' direction in A-form helices.

Another class of experiment widely used in NMR methods correlates resonances based on through-bond scalar couplings. For example, homonuclear COSY (correlated spectroscopy) experiments can be used to establish correlations between protons up to three bonds apart, such as correlations between H5 and H6 in purine bases and between ribose H1'-H2', H2'-H3', H3'-H4', H4'H5' and H4'H5" (129). Long-range correlations can be established using homonuclear TOCSY experiments (total correlated spectroscopy) to correlate, for example, H1' and H3'. Heteronuclear TOCSY experiments, such as an HCP experiment, correlate phosphorous and H4' and H2' protons in addition to the usual H3' and H5' correlations (150). Furthermore, by using a $^{13}C/^{15}N$ labeled RNA sample, an HCCNH TOCSY experiment is able to connect nonexchangeable protons H8 of guanine and H6 of uracil to imino proton resonances through intra-nucleotide bonds (151) and thus establishing the correlation from the imino to aromatic protons. Additionally, direct correlation of resonances between the nucleobase and the ribose sugar can be obtained by HCN triple resonance experiments of isotopically labeled RNA samples (152). The HCN experiment detects the intra-base and sugar to base correlations via stepwise coherence transfer from ^{1}H to ^{13}C to ^{15}N and back through direct spin couplings $^{1}J_{CH}$ and $^{1}J_{CN}$. This provides unambiguous connections of ^{1}H and ^{13}C in the sugar and base of an RNA nucleotide (Figure 1-6). One major breakthrough for the elucidation of base pairing and the complex hydrogen bonding patterns in RNA by NMR spectroscopy is the development of the H_{NN}-COSY experiment (Figure 1-6). It utilizes the sizable $^{2h}J_{NN}$ coupling (on order of 5-7 Hz) to obtain direct correlations between the hydrogen bond donor imino groups of guanine or uridine and the hydrogen bond acceptor N1 and N3 nitrogen atoms of cytosine or adenine, respectively.

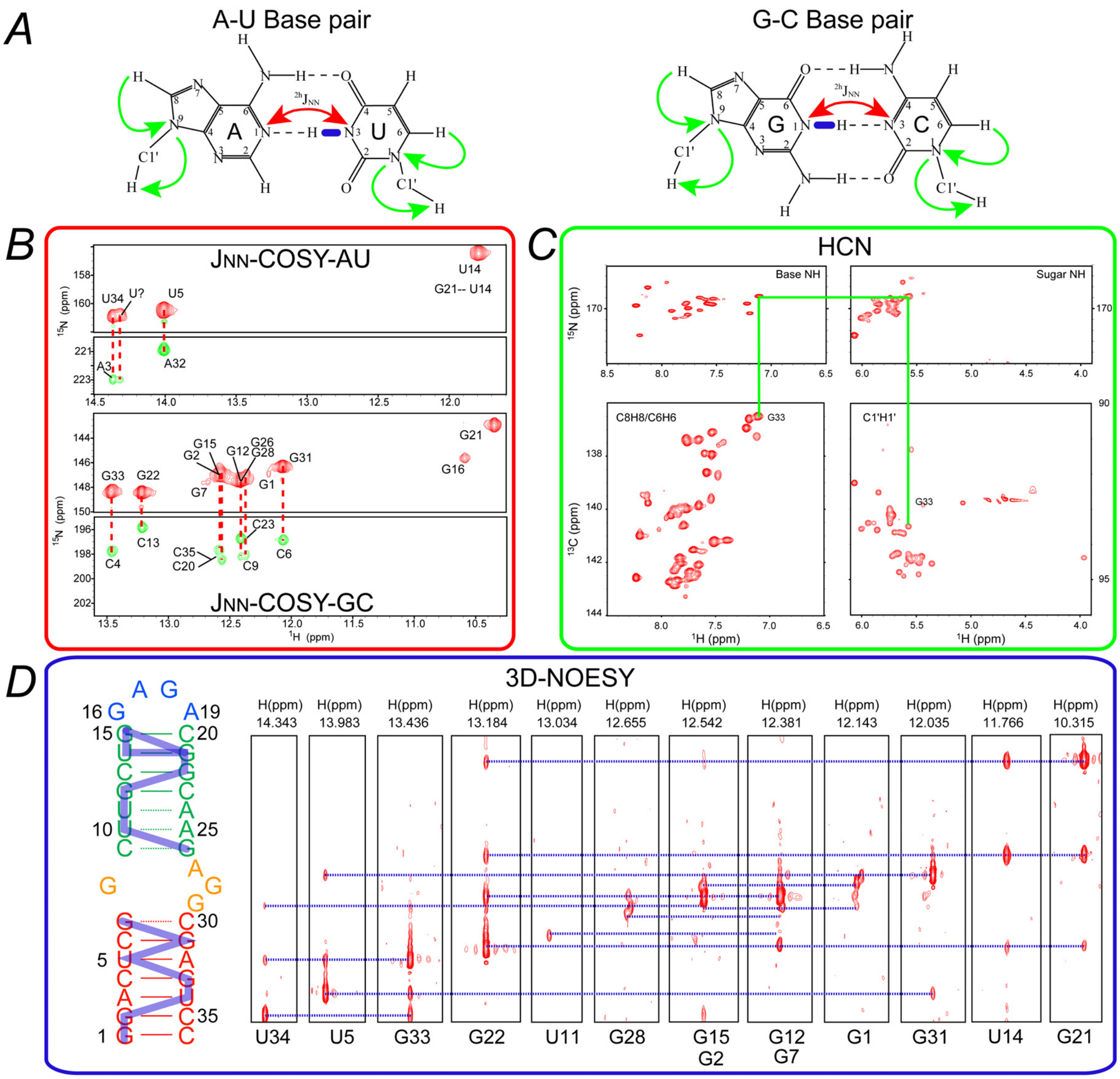

Figure 1-6. NMR assignments of nucleic acids. Shown are assigning of C-H and N-H bonds form the canonic Watson-Crick base pairs by HCN, NOESY, and J_{NN}-COSY experiments.

1.5.3 Characterization of dynamics by spin relaxation and residual dipolar couplings

1.5.3.1 Spin relaxation

The relaxation rates of a given excited nuclear spin back to its thermal equilibrium present dynamical information including both the site-specific internal motions and also the overall motions of a molecule. The most widely used approach to characterize motions at ps-ns timescale by NMR involves the measurement of experimental parameters, such as the longitudinal spin relaxation rate (R_1), the transverse spin relaxation rate (R_2), and the steady state heteronuclear Overhauser enhancement (NOE) of a given nuclei, followed by analysis of the experimental data using the so-called Model Free formalism to extract the motional parameters, such as the amplitude and timescale of the motions (153-155).

The longitudinal spin relaxation rate ($R_1 = 1/T_1$), also known as spin-lattice relaxation rate, describes the recovery rate of the longitudinal magnetization back to the thermal equilibrium. The transverse spin relaxation rate ($R_2 = 1/T_2$), also known as spin-spin relaxation rate, describes the decay rate of transverse magnetization back to zero. While R_1 and R_2 are auto-relaxation rates of a nucleus, the NOE is characterized by the cross-relaxation rate of a dipolar coupled spin pair, in which perturbing the population of one spin leads to the resonance intensity changes of the other spin due to dipole-dipole cross-relaxation. For a spin-1/2 nucleus, such as imino ^{15}N, the relaxation rates are dominated by the dipolar interaction with the bonded proton and chemical shift anisotropy (CSA) of the nitrogen spin. The three relaxation parameters discussed above can be expressed as follows (156):

$$R_1 = \frac{d^2}{4}\left(6J(\omega_H+\omega_N)+3J(\omega_N)+J(\omega_H-\omega_N)\right)+\frac{c^2}{3}J(\omega_N) \tag{1}$$

$$R_2 = \frac{d^2}{8}\left(4J(0)+3J(\omega_N)+6J(\omega_H)+J(\omega_H-\omega_N)+6J(\omega_H+\omega_N)\right)+\frac{c^2}{18}\left(4J(0)+3J(\omega_N)\right)+R_{ex} \tag{2}$$

$$NOE = 1+\frac{d^2\gamma_H}{4R_1\gamma_N}\left(6J(\omega_H+\omega_N)-J(\omega_H-\omega_N)\right) \tag{3}$$

$$d = \mu_o h\gamma_H\gamma_N \Big/ 8\pi^2\left\langle r^3\right\rangle \; ; \; c = \Delta\sigma\omega_N$$

where $J(\omega)$ is the spectral density function, ω is the Larmor frequency in rad s^{-1}, μ_o is the permeability of vacuum, h is Planck's constant, γ_H and γ_N are the gyromagnetic ratios of ^1H and ^{15}N respectively, r is the NH bond length, and $\Delta\sigma$ is the CSA of the ^{15}N.

The spectral density function, $J(\omega)$, contains the desired dynamical information, and $C(t)$ is the Fourier transform of the correlation function that describes the motion of a N-H bond vector due to internal and overall motions. To simplify data interpretation, it is generally assumed that internal and overall motions are not correlated to one another. Thus, the total correlation function ($C(t)$) can be expressed as the product of correlation functions for internal (C_I) and overall (C_O) motions: $C(t)=C_I(t)C_O(t)$. Thus, the Fourier transform leading to the spectral density function can be written as,

$$J(\omega) = 2\int_0^\infty C_O(t)C_I(t)\cos(\omega t)dt \tag{4}$$

a. Isotropic tumbling

If the overall rotational diffusion is isotropic, the overall correlation function can be expressed as:

$$C_O(t) = \frac{1}{5}\left(e^{-t/\tau_m}\right) \tag{5}$$

where τ_m is the overall correlation time.

The internal correlation function is given by,

$$C_I(t) = \langle P_2(\hat{\mu}(0) \cdot \hat{\mu}(t)) \rangle \quad (6)$$

where P_2 (x) is the secondary Legendre polynomial which is given by $(3x^2-1)/2$ and $\hat{\mu}$ is a unit vector describing the bond vector orientation within the molecular frame.

The Lipari and Szabo Model Free formalism is widely used to interpret relaxation parameters in terms of internal motions (157). In Model Free analysis the internal correlation function is modeled by a single exponential function, which has the following form,

$$C_I(t) = S^2 + (1 - S^2) e^{-t/\tau_e} \quad (7)$$

where τ_e is an effective correlation time for internal motions, and S^2 is the order parameter describing the amplitude of internal motions, which ranges between 0 to 1 for completely isotropic to rigid internal motions. An expression for the special density function can be obtained by taking the Fourier transform of Equations (4), (5), (7),

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_m}{1+(\omega\tau_m)^2} + \frac{(1-S^2)\tau}{1+(\omega\tau)^2}\right] \quad (8)$$

where $\tau = \tau_m\tau_e/(\tau_m + \tau_e)$. The parameters, S^2, τ_m, and τ_e, can be determined based on measurements of R_1, R_2, and NOE for a variety of nuclei in a given target. More complex internal motions can also be accommodated in the Model Free approach. In the extended Model Free formalism introduced by Clore and co-workers, two distinct correlation times can be used to model internal motions (158),

$$C_I(t) = S^2 + (1 - S_f^2) e^{-t/\tau_f} + (S_f^2 - S^2) e^{-t/\tau_s} \quad (9)$$

in which $S^2 = S^2_f S^2_s$ and the S_f and τ_f and S_s and τ_s describe “fast” and “slow” motions of the bond vector, respectively. The spectral density function can be expressed as,

$$J(\omega)=\frac{2}{5}\left[\frac{S^2\tau_m}{1+(\omega\tau_m)^2}+\frac{\left(1-S_f^2\right)\tau_{f}}{1+\left(\omega\tau_{f}\right)^2}+\frac{\left(S_f^2-S^2\right)\tau_{s}}{1+\left(\omega\tau_{s}\right)^2}\right]$$
$$\tau_{f,s}=\frac{\tau_{f,s}\tau_m}{\tau_{f,s}+\tau_m} \qquad (10)$$

For very fast internal motions, Equation (10) can be simplified as,

$$J(\omega)=\frac{2}{5}S_f^2\left[\frac{S_s^2\tau_m}{1+(\omega\tau_m)^2}+\frac{\left(1-S_s^2\right)\tau_{s}}{1+\left(\omega\tau_{s}\right)^2}\right]$$
$$\tau_{s}=\frac{\tau_s\tau_m}{\tau_s+\tau_m} \qquad (11)$$

b. Anisotropic tumbling

For anisotropic overall motions, the overall correlation function is given by (157)

$$C_O(t)=\frac{1}{5}Ae^{-t/\tau_1}+\frac{1}{5}(1-A)e^{-t/\tau_2} \qquad (12)$$

where A, τ_1 and τ_2 are adjustable parameters for fitting the experimental relaxation data. As described by Lipari and Szabo (157), the total correlation function for anisotropic tumbling is,

$$C(t)=\frac{1}{5}\left(Ae^{-t/\tau_1}+(1-A)e^{-t/\tau_2}\right)\left(S^2+\left(1-S^2\right)e^{-t/\tau_e}\right) \qquad (13)$$

where τ_e is an effective correlation time for internal motions, and S^2 is the spin relaxation order parameter describing the amplitude of internal motions. For axially symmetric anisotropic tumbling, the overall rotational diffusion tensor is characterized by

$D_\perp = D_x = D_y$ and $D_{//} = D_z$. According to the extended Model Free approach (158), the spectral density can be written as,

$$J(\omega) = \frac{2}{5} S_f^2 \sum_{j=1}^{3} A_j \left[\frac{S_s^2 \tau_j}{1+(\omega\tau_j)^2} + \frac{(1-S_s^2)\tau'_j}{1+(\omega\tau'_j)^2} \right] \qquad (14)$$

$$\tau'_j = \frac{\tau_j \tau_s}{\tau_j + \tau_s}$$

where $A_1 = \left(\frac{3\cos^2\theta - 1}{2}\right)^2$, $A_2 = 3\sin^2\theta\cos^2\theta$, $A_3 = \frac{3}{4}\sin^4\theta$, $\tau_1^{-1} = 6D_\perp$, $\tau_2^{-1} = 5D_\perp + D_{//}$, $\tau_3^{-1} = 2D_\perp + 4D_{//}$, and θ is the average angle between the bond vector and principle axis (D_{zz}) of the rotational diffusion tensor.

c. Experimental measurement of ^{15}N relaxation

The imino ^{15}N relaxation parameters on RNA, R_1, R_2, and ^{1}H–^{15}N NOE, can be measured using pulse sequence shown in Figure 1-7. Both R_1 and R_2 pulse sequences start with a refocused INEPT transfer, preparing nitrogen N_x magnetization. For R_1 experiments, a 90^o pulse converts N_x into N_z prior to the relaxation period, while the magnetization is kept along N_x in R_2 experiments. A selective 90^o pulse is applied in order to position water magnetization along +z. The relaxation period is followed by t_1 evolution period, reverse polarization transfer, and detection of imino proton magnetization. During the relaxation period, selective 180° shaped pulses centered on the imino proton are applied to suppress the effects of ^{1}H –^{15}N dipolar cross-relaxation and ^{15}N CSA/^{1}H –^{15}N dipolar relaxation interference. During the relaxation period in the R_2 experiment, the Carr-Purcell-Meiboom-Gill (CPMG) train of 180° pulses is applied in order to suppress chemical exchange (R_{ex}), with a recently developed $[0013]^N$ phase cycling scheme in order to suppress artifacts arising from off-resonance effects (159). In addition, low power CPMG train pulses are used to avoid excitation of the hydrogen bond acceptor nitrogen N3 (cytosine) and N1 (adenine). For the heteronuclear ^{1}H–^{15}N NOE experiment, two experiments are carried out with and without proton saturation, and NOE enhancements can be measured as the ratio of resonance intensities with and without proton saturation. The saturation of proton leads to a change in the nitrogen magnetization due to dipolar cross-relaxation, which is achieved as a series of 120^o pulses, and the water suppression is achieved using water flip-back pulses (160) (161).

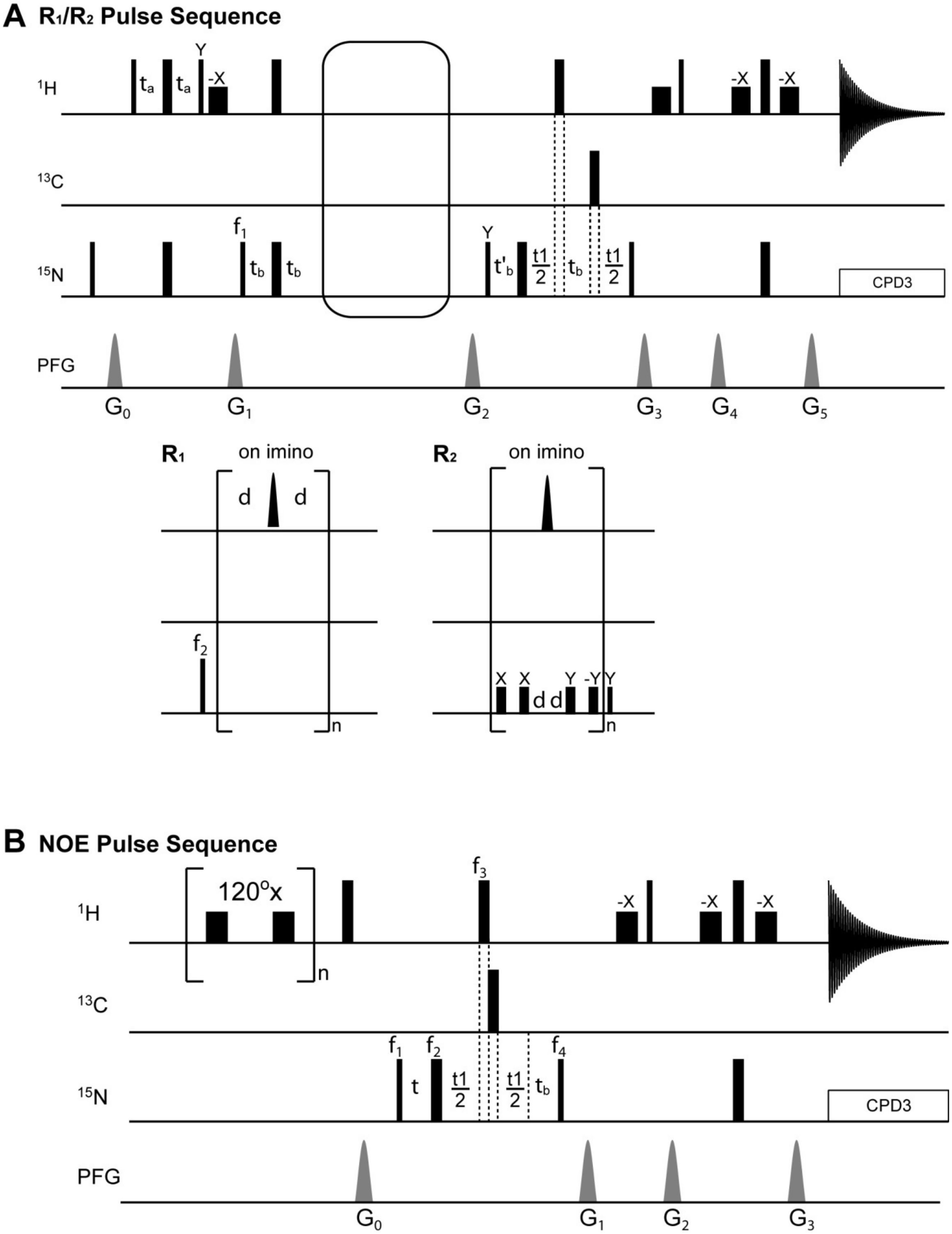

Figure 1-7. Measurement of ^{15}N spin relaxation in RNA by NMR. Shown are 2D proton detected pulse sequences for measuring (A) ^{15}N R_1 and R_2 and (B)^{1}H-^{15}N NOE.

1.5.3.2 Residual dipolar couplings

Nuclear dipole-dipole interactions arise due to the modulation of the local field of a nucleus by the nuclear magnetic flux arising from a neighboring nucleus. The expression for the heteronuclear dipolar couplings between two spin ½ nuclei (i and j) is given by,

$$D_{ij} = -\frac{\gamma_i \gamma_j \mu_o h}{8\pi^3 \langle r_{ij}^3 \rangle} \left\langle \frac{(3\cos^2\theta_{ij} - 1)}{2} \right\rangle \tag{15}$$

where γ is the gyromagnetic ratios of nuclei, μ_o is the magnetic permittivity of vacuum, h is Plank's constant, r_{ij} is the distance between the two nuclei, and θ_{ij} is the angle between the vector connecting the two nuclei and the applied magnetic field. The angular term within brackets is the time average of the angles sampled by the vector due to internal and overall motions over timescales faster than the inverse of the dipolar interaction (typically < ms). Dipolar couplings are normally not observed under solution conditions, since the angular term averages to zero due to overall tumbling, which is random due to isotropic Brownian motion (Figure 1-8A). However, a degree of alignment can be imparted on the solute of interest either spontaneously using the magnetic field or by using an ordering medium (162). Under these conditions, the angular term no longer averages to zero. As is the case for through-bond scalar couplings (J), through-space dipolar couplings (D) effectively increase or decrease the average magnetic field at a given nucleus, resulting in the splitting of resonances. Dipolar couplings can thereby be measured as contributions to scalar couplings (J) that are observed under conditions of molecular alignment (J+D). The most commonly used media for nucleic acids is filamentous phage due to its low nematic threshold concentration and negative charge, which reduces the possibility of interaction between the alignment media and the nucleic acids. The optimum level of alignment is on the order of 10^{-3}, which means approximately 1 in 1,000 RNA molecules are completely aligned. Such levels of alignment can be achieved for small RNA constructs (<30 nt) using phage concentration of ~18-25 mg/ml. Under these conditions, RDCs can be measured with an optimal magnitude/precision ratio and with minimal sacrifice in spectral resolution. One bond C-H and N-H RDCs in base and ribose moieties of RNA are the most commonly targeted interactions due to their favorable size, smaller one bond C-C and C-N as well as two and three bond RDCs can be also measured (144,163) (Figure 1-8C and D).

The measurement of RDCs in partially aligned systems provides long-range orientational constraints that are highly complementary to the NOE derived short range proton-proton distance constraints. These long-range constraints are particularly important in defining the global topology of nucleic acids. In addition, RDCs can provide unique information about internal motions occurring over a wide window of biologically relevant timescales (<ms) (Figure 1-10B).

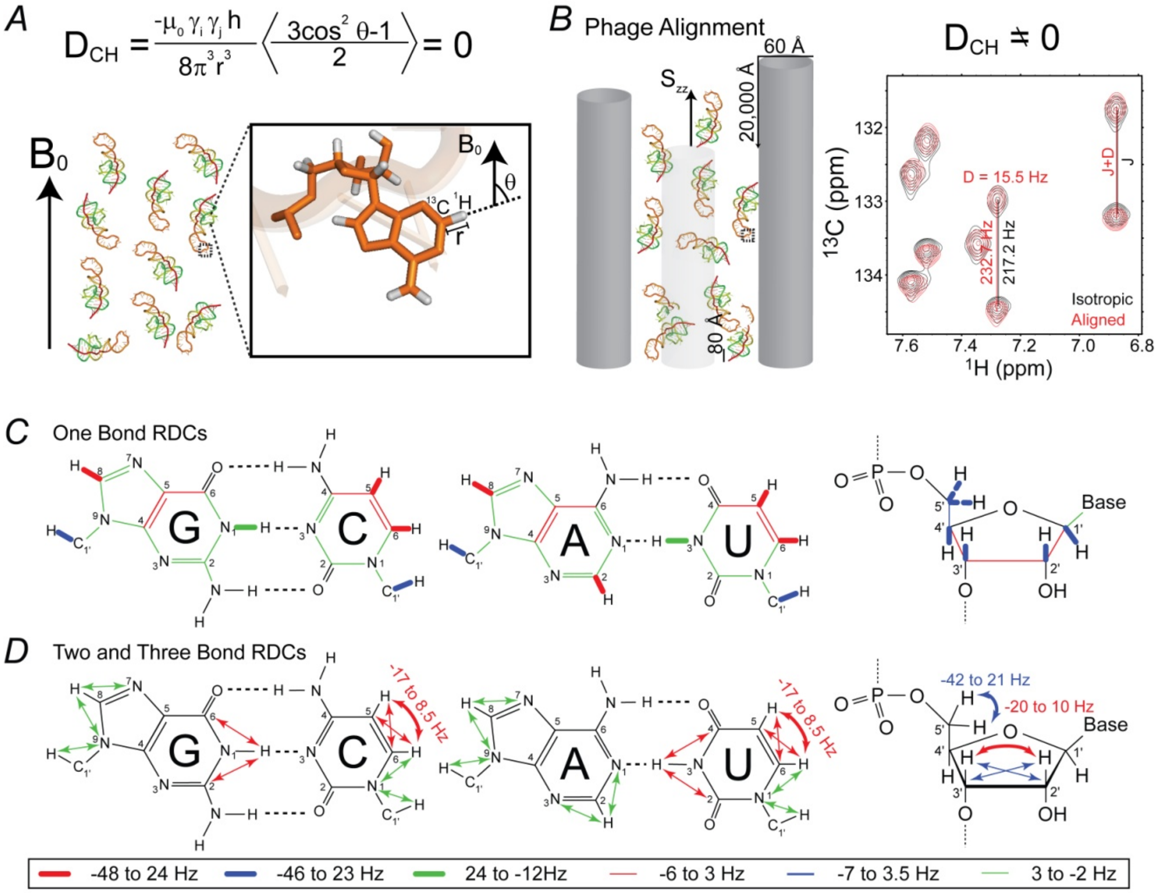

Figure 1-8. Typical measurement of RNA RDCs. (A) No aligment (B) Partial alignment of RNA using ordering media Pf1 phage. (C-D) RDCs measured in base and sugar of RNA.

The analysis of RDCs requires the interpretation of the angular term in Equation 11. The angular term can be expressed in terms of the time dependent orientation, *Φ,* of an internuclear vector in the molecular frame and five elements of an order tensor ($S_{k,l}$) which describe overall alignment of the molecule,

$$\left\langle \frac{(3\cos^2\theta_{ij}-1)}{2} \right\rangle = \sum_{k,l=x,y,z} S_{k,l}\cos\phi_k^{ij}\cos\phi_l^{ij} \qquad (16)$$

and

$$S_{k,l} = \begin{pmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{pmatrix} \tag{17}$$

Since the order matrix $S_{k,l}$ is symmetric ($S_{k,l} = S_{l,k}$) and traceless ($S_{xx} + S_{yy} + S_{zz} = 0$), there are total of five independent parameters. The two principal order parameters define the degree ($\vartheta = \sqrt{\frac{2}{3}\left(S_{xx}^2 + S_{yy}^2 + S_{zz}^2\right)}$, $|S_{zz}| \geq |S_{yy}| \geq |S_{xx}|$) and asymmetry ($\eta = \frac{|S_{yy} - S_{xx}|}{S_{zz}}$) of molecular alignment; two angular parameters define the average orientation of the magnetic field relative to the chiral molecular frame (the principal direction, S_{zz}), and a third angular parameter defines the orientation of an orthogonal principal axis (S_{yy}) that specifies the asymmetry of alignment. The five order tensor elements can be determined experimentally provided measurement of five or more independent RDCs in a fragment with known local structure.

One of the first approaches for interpreting RDCs in terms of structure and dynamics involves determining order tensors for fragments, and this proves to be of general utility in studies of RNA (146,164,165). RNAs, by their own naturally properties, can be decomposed into sub-structures consisting of locally stable A-form helices, the most abundant RNA secondary element (143,166). The determination of order tensors for individual helical fragments provide a natural approach for determining their relative orientation and dynamics (143,144). The orientation and dynamics of A-form helical domains is an important feature of RNA architecture that goes through mechanistically important changes during folding, recognition and catalysis(136,167,168). It is also a feature of RNA architecture that is prone to artificial distortions from crystal packing forces.

The relative orientation of helices can be determined by superimposing their order tensor frames (Figure 1-9). The latter step insists that the helical domains share a common average view of the magnetic field direction when assembled into a proper structure. Difficulties arise due to the fact that RDCs are degenerate to 180° rotations around the principal directions of the order tensor (S_{xx}, S_{yy}, and S_{zz}). This results in 4^{n-1} fold degeneracy in orienting *n* fragments. This degeneracy can be overcome either by measuring RDCs under at least two different alignments or, more typically in nucleic acids, by incorporating additional experimental and non-experimental restraints.

The two principal order tensor parameters (ϑ and η) obtained for each helical domain can be compared to obtain information about relative helix motions over sub-millisecond timescales (Figure 1-9). While helices will report identical parameters when they are rigid relative to one another, inter-helix motions can lead to differences. Specifically, the degree of order for a given helix (ϑ) will be attenuated relative to the value observed for a helix dominating alignment, with the degree of attenuation increasing with the amplitude of motions. The ratio of alignment order, defined as the internal generalized degree of order ($\vartheta_{int} = \vartheta_i/\vartheta_j$; $\vartheta_i<\vartheta_j$)) then provides a measurement of motional amplitudes, with $\vartheta_{int} = 1$ corresponding to perfect rigidity and $\vartheta_{int} = 0$ to maximum motions. Although often difficult to determine reliably, the asymmetry parameter (η) can provide insight into the directionality of motions with isotropic directionless motions having a smaller effect compared to anisotropic directional motions.

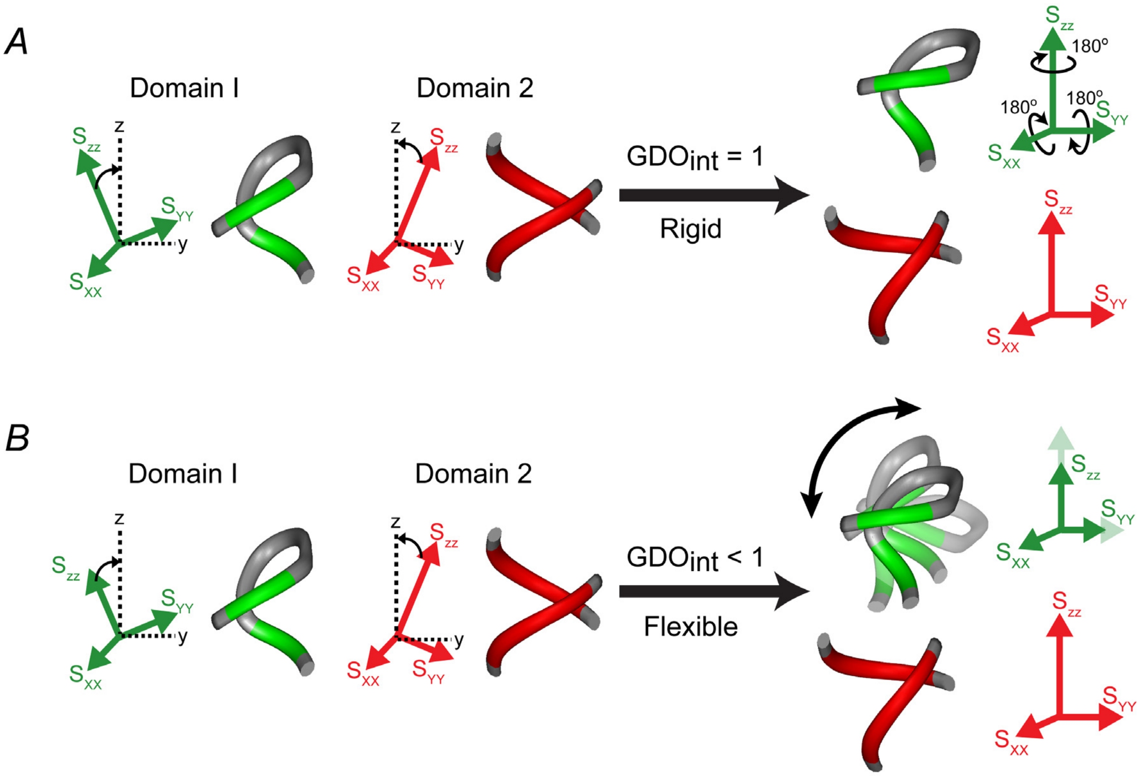

Figure 1-9. Structure and dynamics determination by RDC analysis. Shown are (A) structural and (B) dynamical characterization of domains.

1.6 Methods to characterize RNA-ligand interactions

Metal cations provide a means for neutralizing undesirable charge repulsion between negatively charged regions in RNA, permitting complex compact structures to form. Metal cations also play a direct role in ribozyme catalysis: tRNA^{Phe} cleaves itself in the presence of Pb^{2+}, and the HDV ribozyme, as well as the larger Group I and Group II introns and RNase P, appear to be dependent upon divalent metal ions for catalysis (169-172). Most RNA molecules contain specific sites that can strongly associate with mono- and divalent metal ions. There are three classes of RNA-metal interactions, diffusely bound and site-bound (173-175). Diffusely bound ions are often referred to as a “delocalized cation atmosphere”, and site bound metal ions usually make two or more direct contacts to functional groups in the RNA either through outer or inner sphere interactions (176-178).

Three NMR approaches are often applied to characterize RNA and metal interactions: probing chemical shift changes, detecting intermolecular NOEs, and measuring paramagnetic relaxation enhancement (179,180). Tracking chemical shift changes of resonances of traditional ^{1}H, ^{15}N, or ^{13}C nuclei and the ^{31}P from the negatively charged RNA backbone is the most direct way to identify metal-sensitive sites, which are affected by the deshielding effects of the ion binding or by the local or global structural transition induced upon ion binding. In addition, by plotting the chemical shift changes as a function of metal concentration, the apparent metal binding affinity of individual residues can be obtained assuming a two-state binding model.

Another NMR approach, paramagnetic resonance enhancement (PRE), can provide detailed metal localization information of an RNA-metal complex (181,182). In this method, a paramagnetic ion, like Mn^{2+}, causes line broadening of NMR resonances within close proximity, approximately, ~1 nm. This effect is due to the increase of the relaxation rate of the nuclear spins. The extent of this effect is dependent on the degree of occupancy of the metal binding site and is proportional to r^{-6}, where r is the distance between the observed nucleus and the paramagnetic ion. In the presence of a very low concentration of paramagnetic ion, the ions only bind to the specific region(s) of an RNA molecule with high affinity, thereby the metal binding site(s).

It is often of interest to determine the interaction sites within a complex of RNA and its ligand, such as protein and other RNA molecules (183)(184)(185). That information could provide useful insight and give light to the interpretation of biological studies. It is not always easy to solve a complete RNA-ligand structure under physiological conditions, and it is not efficient or relevant to do mutation studies. Current NMR strategies contain straightforward methods to obtain the site specific information of the inter-surface of a large RNA-ligand (186). Especially in RNA-protein complexes, probing the chemical shift changes identifies the individual residues which may be involved in the fast or intermediate exchange in on- or off-bound states (183). Also, by recording the imino proton resonances of RNA upon the gradual increase of protein concentration, additional signals indicate the special interface resolved either by stabilization of the RNA structure or by protection from exchange by direct RNA-protein interactions.

Moreover, the cross saturation method can identify the residues at the interface between protein and RNA by observing the reductions of the peak intensities in the HSQC spectra measured by TROSY coherence transfer (186). In this method, ^{1}H resonances from a proton-rich component of the complex are selectively saturated by irradiation. By spin diffusion, this saturation is transferred to the uniformly ^{2}H and ^{15}N labeled second component of the complex, thereby leading to a loss of the signal intensities of its NMR spectra resonances. The peaks that are affected the most are close to the interface of the complex. This allows for direct identification of residues that are part of the interaction surface.

1.7 References

1. Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C. *et al.* (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, **220**, 868-871.
2. Reeves, J.D. and Doms, R.W. (2002) Human immunodeficiency virus type 2. *J Gen Virol*, **83**, 1253-1265.
3. Gao, F., Yue, L., White, A.T., Pappas, P.G., Barchue, J., Hanson, A.P., Greene, B.M., Sharp, P.M., Shaw, G.M. and Hahn, B.H. (1992) Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. *Nature*, **358**, 495-499.
4. Weissbrich, B., Heinkelein, M. and Jassoy, C. (2002), *Advances in Virus Research*. Academic Press, Vol. Volume 58, pp. 157-202.
5. De Clercq, E. (1994) HIV resistance to reverse transcriptase inhibitors. *Biochemical Pharmacology*, **47**, 155-169.
6. Wang, Z., Bennett, E.M., Wilson, D.J., Salomon, C. and Vince, R. (2007) Rationally Designed Dual Inhibitors of HIV Reverse Transcriptase and Integrase. *J. Med. Chem.*, **50**, 3416-3419.
7. Klimas, N., Koneru, A.O.B. and Fletcher, M.A. (2008) Overview of HIV. *Psychosom Med*, **70**, 523-530.
8. Tor, Y. (2003) Targeting RNA with small molecules. *Chembiochem*, **4**, 998-1007.
9. Vicens, Q. and Westhof, E. (2003) RNA as a drug target: the case of aminoglycosides. *Chembiochem*, **4**, 1018-1023.
10. Ennifar, E., Paillart, J.C., Bodlenner, A., Walter, P., Weibel, J.M., Aubertin, A.M., Pale, P., Dumas, P. and Marquet, R. (2006) Targeting the dimerization initiation site of HIV-1 RNA with aminoglycosides: from crystal to cell. *Nucleic Acids Res*, **34**, 2328-2339.
11. Shubsda, M.F., Paoletti, A.C., Hudson, B.S. and Borer, P.N. (2002) Affinities of packaging domain loops in HIV-1 RNA for the nucleocapsid protein. *Biochemistry*, **41**, 5276-5282.
12. Hagan, N.A. and Fabris, D. (2007) Dissecting the protein-RNA and RNA-RNA interactions in the nucleocapsid-mediated dimerization and isomerization of HIV-1 stemloop 1. *J Mol Biol*, **365**, 396-410.
13. de Rocquigny, H., Shvadchak, V., Avilov, S., Dong, C.Z., Dietrich, U., Darlix, J.L. and Mely, Y. (2008) Targeting the viral nucleocapsid protein in anti-HIV-1 therapy. *Mini Rev Med Chem*, **8**, 24-35.
14. Varmus, H.E. (1982) Form and Function of Retroviral Proviruses. *Science*, **216**, 812-820.
15. Henderson, L.E., Bowers, M.A., Sowder, R.C., 2nd, Serabyn, S.A., Johnson, D.G., Bess, J.W., Jr., Arthur, L.O., Bryant, D.K. and Fenselau, C. (1992) Gag proteins of the highly replicative MN strain of human immunodeficiency virus type 1: posttranslational modifications, proteolytic processings, and complete amino acid sequences. *J. Virol.*, **66**, 1856-1865.
16. Bender, W. and Davidson, N. (1976) Mapping of poly(A) sequences in the electron microscope reveals unusual structure of type C oncovirus RNA molecules. *Cell*, **7**, 595 - 607.

17. Robinson, W.S., Robinson, H.L. and Duesberg, P.H. (1967) Tumor virus RNA's. *Proc Natl Acad Sci U S A*, **58**, 825 - 834.
18. Robinson, W.S., Robinson, H.L. and Duesberg, P.H. (1968) Tumor virus RNA's. *Proc Natl Acad Sci U S A*, **58**, 825 - 834.
19. Kung, H.J., Bailey, J.M., Davidson, N., Nicolson, M.O. and McAllister, R.M. (1975) Structure, subunit composition, and molecular weight of RD-114 RNA. *J Virol*, **16**, 397 - 411.
20. Bender, W., Chien, Y.H., Chattopadhyay, S., Vogt, P.K., Gardner, M.B. and Davidson, N. (1978) High-molecular-weight RNAs of AKR, NZB, and wild mouse viruses and avian reticuloendotheliosis virus all have similar dimer structures. *J Virol*, **25**, 888 - 896.
21. Clever, J.L. and Parslow, T.G. (1997) Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J Virol*, **71**, 3407-3414.
22. Murti, K.G., Bondurant, M. and Tereba, A. (1981) Secondary structural features in the 70S RNAs of Moloney murine leukemia and Rous sarcoma viruses as observed by electron microscopy. *J Virol*, **37**, 411 - 419.
23. Paillart, J.C., Marquet, R., Skripkin, E., Ehresmann, B. and Ehresmann, C. (1994) Mutational analysis of the bipartite dimer linkage structure of human immunodeficiency virus type 1 genomic RNA. *J Biol Chem*, **269**, 27486-27493.
24. Paillart, J.C., Marquet, R., Skripkin, E., Ehresmann, C. and Ehresmann, B. (1996) Dimerization of retroviral genomic RNAs: structural and functional implications. *Biochimie*, **78**, 639-653.
25. Paillart, J.C., Shehu-Xhilaga, M., Marquet, R. and Mak, J. (2004) Dimerization of retroviral RNA genomes: an inseparable pair. *Nat Rev Microbiol*, **2**, 461 - 472.
26. Baba, S., Takahashi, K., Nomura, Y., Noguchi, S., Koyanagi, Y., Yamamoto, N., Takaku, H. and Kawai, G. (2001) Conformational change of dimerization initiation site of HIV-1 genomic RNA by NCp7 or heat treatment. *Nucleic Acids Res Suppl*, 155-156.
27. Takahashi, K., Baba, S., Hayashi, Y., Koyanagi, Y., Yamamoto, N., Takaku, H. and Kawai, G. (2000) NMR analysis of intra- and inter-molecular stems in the dimerization initiation site of the HIV-1 genome. *J Biochem*, **127**, 681-686.
28. Fu, W., Gorelick, R.J. and Rein, A. (1994) Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. *J Virol*, **68**, 5013 - 5018.
29. Fu, W. and Rein, A. (1993) Maturation of dimeric viral RNA of Moloney murine leukemia virus. *J Virol*, **67**, 5443 - 5449.
30. Sheng, N. and Erickson-Viitanen, S. (1994) Cleavage of p15 protein in vitro by human immunodeficiency virus type 1 protease is RNA dependent. *J. Virol.*, **68**, 6207-6214.
31. Liang, C., Rong, L., Quan, Y., Laughrea, M., Kleiman, L. and Wainberg, M.A. (1999) Mutations within four distinct gag proteins are required to restore replication of human immunodeficiency virus type 1 after deletion mutagenesis within the dimerization initiation site. *J Virol*, **73**, 7014-7020.
32. Muriaux, D., Mirro, J., Harvin, D. and Rein, A. (2001) RNA is a structural element in retrovirus particles. *Proc Natl Acad Sci U S A*, **98**, 5246 - 5251.

33. Feng, Y.X., Campbell, S., Harvin, D., Ehresmann, B., Ehresmann, C. and Rein, A. (1999) The human immunodeficiency virus type 1 Gag polyprotein has nucleic acid chaperone activity: possible role in dimerization of genomic RNA and placement of tRNA on the primer binding site. *J Virol*, **73**, 4251-4256.
34. Shehu-Xhilaga, M., Crowe, S.M. and Mak, J. (2001) Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. *J Virol*, **75**, 1834 - 1841.
35. Shehu-Xhilaga, M., Hill, M., Marshall, J.A., Kappes, J., Crowe, S.M. and Mak, J. (2002) The conformation of the mature dimeric human immunodeficiency virus type 1 RNA genome requires packaging of pol protein. *J Virol*, **76**, 4331-4340.
36. Shehu-Xhilaga, M., Kraeusslich, H.G., Pettit, S., Swanstrom, R., Lee, J.Y., Marshall, J.A., Crowe, S.M. and Mak, J. (2001) Proteolytic processing of the p2/nucleocapsid cleavage site is critical for human immunodeficiency virus type 1 RNA dimer maturation. *J Virol*, **75**, 9156-9164.
37. Paillart, J.C., Skripkin, E., Ehresmann, B., Ehresmann, C. and Marquet, R. (1996) A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc Natl Acad Sci U S A*, **93**, 5572-5577.
38. Fosse, P., Motte, N., Roumier, A., Gabus, C., Muriaux, D., Darlix, J.L. and Paoletti, J. (1996) A short autocomplementary sequence plays an essential role in avian sarcoma-leukosis virus RNA dimerization. *Biochemistry*, **35**, 16601 - 16609.
39. Berkhout, B. and van Wamel, J.L. (1996) Role of the DIS hairpin in replication of human immunodeficiency virus type 1. *J Virol*, **70**, 6723-6732.
40. Takahashi, K.I., Baba, S., Chattopadhyay, P., Koyanagi, Y., Yamamoto, N., Takaku, H. and Kawai, G. (2000) Structural requirement for the two-step dimerization of human immunodeficiency virus type 1 genome. *Rna*, **6**, 96-102.
41. Darlix, J.L., Gabus, C., Nugeyre, M.T., Clavel, F. and Barre-Sinoussi, F. (1990) Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J Mol Biol*, **216**, 689 - 699.
42. Weiss, S., Konig, B., Morikawa, Y. and Jones, I. (1992) Recombinant HIV-1 nucleocapsid protein p15 produced as a fusion protein with glutathione S-transferase in Escherichia coli mediates dimerization and enhances reverse transcription of retroviral RNA. *Gene*, **121**, 203-212.
43. Sakaguchi, K., Zambrano, N., Baldwin, E.T., Shapiro, B.A., Erickson, J.W., Omichinski, J.G., Clore, G.M., Gronenborn, A.M. and Appella, E. (1993) Identification of a binding site for the human immunodeficiency virus type 1 nucleocapsid protein. *Proc Natl Acad Sci U S A*, **90**, 5219 - 5223.
44. De Rocquigny, H., Gabus, C., Vincent, A., Fournie-Zaluski, M.C., Roques, B. and Darlix, J.L. (1992) Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers. *Proc Natl Acad Sci U S A*, **89**, 6472-6476.
45. Huthoff, H., Das, A.T., Vink, M., Klaver, B., Zorgdrager, F., Cornelissen, M. and Berkhout, B. (2004) A human immunodeficiency virus type 1-infected individual with low viral load harbors a virus variant that exhibits an in vitro RNA dimerization defect. *J Virol*, **78**, 4907-4913.

46. Sakuragi, J.-i., Sakuragi, S. and Shioda, T. (2007) Minimal Region Sufficient for Genome Dimerization in the Human Immunodeficiency Virus Type 1 Virion and Its Potential Roles in the Early Stages of Viral Replication. *J. Virol.*, **81**, 7985-7992.
47. Balakrishnan, M., Roques, B.P., Fay, P.J. and Bambara, R.A. (2003) Template dimerization promotes an acceptor invasion-induced transfer mechanism during human immunodeficiency virus type 1 minus-strand synthesis. *J Virol*, **77**, 4710-4721.
48. Balakrishnan, M., Fay, P.J. and Bambara, R.A. (2001) The kissing hairpin sequence promotes recombination within the HIV-I 5' leader region. *J Biol Chem*, **276**, 36482-36492.
49. Gotte, M., Li, X. and Wainberg, M.A. (1999) HIV-1 reverse transcription: a brief overview focused on structure-function relationships among molecules involved in initiation of the reaction. *Arch Biochem Biophys*, **365**, 199-210.
50. Berkhout, B., Das, A.T. and van Wamel, J.L. (1998) The native structure of the human immunodeficiency virus type 1 RNA genome is required for the first strand transfer of reverse transcription. *Virology*, **249**, 211 - 218.
51. Shen, N., Jette, L., Liang, C., Wainberg, M.A. and Laughrea, M. (2000) Impact of human immunodeficiency virus type 1 RNA dimerization on viral infectivity and of stem-loop B on RNA dimerization and reverse transcription and dissociation of dimerization from packaging. *J Virol*, **74**, 5729-5735.
52. Sakuragi, J., Ueda, S., Iwamoto, A. and Shioda, T. (2003) Possible role of dimerization in human immunodeficiency virus type 1 genome RNA packaging. *J Virol*, **77**, 4060-4069.
53. Sakuragi, J., Iwamoto, A. and Shioda, T. (2002) Dissociation of genome dimerization from packaging functions and virion maturation of human immunodeficiency virus type 1. *J Virol*, **76**, 959-967.
54. De Guzman, R.N., Wu, Z.R., Stalling, C.C., Pappalardo, L., Borer, P.N. and Summers, M.F. (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science*, **279**, 384 - 388.
55. Amarasinghe, G.K., De Guzman, R.N., Turner, R.B., Chancellor, K.J., Wu, Z.R. and Summers, M.F. (2000) NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *J Mol Biol*, **301**, 491-511.
56. Amarasinghe, G.K., Zhou, J., Miskimon, M., Chancellor, K.J., McDonald, J.A., Matthews, A.G., Miller, R.R., Rouse, M.D. and Summers, M.F. (2001) Stem-loop SL4 of the HIV-1 psi RNA packaging signal exhibits weak affinity for the nucleocapsid protein. structural studies and implications for genome recognition. *J Mol Biol*, **314**, 961-970.
57. D'Souza, V. and Summers, M.F. (2005) How retroviruses select their genomes. *Nat Rev Microbiol*, **3**, 643-655.
58. Houzet, L., Paillart, J.C., Smagulova, F., Maurel, S., Morichaud, Z., Marquet, R. and Mougel, M. (2007) HIV controls the selective packaging of genomic, spliced viral and cellular RNAs into virions through different mechanisms. *Nucleic Acids Res*, **35**, 2695-2704.

59. Kim, H.J., Lee, K. and O'Rear, J.J. (1994) A short sequence upstream of the 5' major splice site is important for encapsidation of HIV-1 genomic RNA. *Virology*, **198**, 336 - 340.

60. Grigorov, B., Decimo, D., Smagulova, F., Pechoux, C., Mougel, M., Muriaux, D. and Darlix, J.L. (2007) Intracellular HIV-1 Gag localization is impaired by mutations in the nucleocapsid zinc fingers. *Retrovirology*, **4**, 54.

61. Ooms, M., Huthoff, H., Russell, R., Liang, C. and Berkhout, B. (2004) A riboswitch regulates RNA dimerization and packaging in human immunodeficiency virus type 1 virions. *J Virol*, **78**, 10814-10819.

62. Huthoff, H. and Berkhout, B. (2001) Two alternating structures of the HIV-1 leader RNA. *Rna*, **7**, 143-157.

63. Abbink, T.E., Ooms, M., Haasnoot, P.C. and Berkhout, B. (2005) The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. *Biochemistry*, **44**, 9058-9066.

64. Gao, F., Yue, L., Craig, S., Thornton, C.L., Robertson, D.L., McCutchan, F.E., Bradac, J.A., Sharp, P.M. and Hahn, B.H. (1994) Genetic variation of HIV type 1 in four World Health Organization-sponsored vaccine evaluation sites: generation of functional envelope (glycoprotein 160) clones representative of sequence subtypes A, B, C, and E. WHO Network for HIV Isolation and Characterization. *AIDS Res Hum Retroviruses*, **10**, 1359-1368.

65. Gao, F., Yue, L., Hill, S.C., Robertson, D.L., Graves, A.H., Saag, M.S., Shaw, G.M., Sharp, P.M. and Hahn, B.H. (1994) HIV-1 sequence subtype D in the United States. *AIDS Res Hum Retroviruses*, **10**, 625-627.

66. Gao, F., Robertson, D.L., Morrison, S.G., Hui, H., Craig, S., Decker, J., Fultz, P.N., Girard, M., Shaw, G.M., Hahn, B.H. *et al.* (1996) The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J Virol*, **70**, 7013-7029.

67. Gao, F., Morrison, S.G., Robertson, D.L., Thornton, C.L., Craig, S., Karlsson, G., Sodroski, J., Morgado, M., Galvao-Castro, B., von Briesen, H. *et al.* (1996) Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. The WHO and NIAID Networks for HIV Isolation and Characterization. *J Virol*, **70**, 1651-1667.

68. Gurtler, L.G., Hauser, P.H., Eberle, J., von Brunn, A., Knapp, S., Zekeng, L., Tsague, J.M. and Kaptue, L. (1994) A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. *J Virol*, **68**, 1581-1585.

69. Marquet, R., Baudin, F., Gabus, C., Darlix, J.L., Mougel, M., Ehresmann, C. and Ehresmann, B. (1991) Dimerization of human immunodeficiency virus (type 1) RNA: stimulation by cations and possible mechanism. *Nucleic Acids Res*, **19**, 2349 - 2357.

70. Laughrea, M. and Jette, L. (1994) A 19-nucleotide sequence upstream of the 5' major splice donor is part of the dimerization domain of human immunodeficiency virus 1 genomic RNA. *Biochemistry*, **33**, 13464 - 13474.

71. Skripkin, E., Paillart, J.C., Marquet, R., Ehresmann, B. and Ehresmann, C. (1994) Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *Proc Natl Acad Sci U S A*, **91**, 4945 - 4949.

72. Clever, J.L., Wong, M.L. and Parslow, T.G. (1996) Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA. *J Virol*, **70**, 5902-5908.
73. Lodmell, J.S., Ehresmann, C., Ehresmann, B. and Marquet, R. (2001) Structure and dimerization of HIV-1 kissing loop aptamers. *J Mol Biol*, **311**, 475-490.
74. Shen, N., Jette, L., Wainberg, M.A. and Laughrea, M. (2001) Role of stem B, loop B, and nucleotides next to the primer binding site and the kissing-loop domain in human immunodeficiency virus type 1 replication and genomic-RNA dimerization. *J Virol*, **75**, 10543 - 10549.
75. Clever, J., Sassetti, C. and Parslow, T.G. (1995) RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *J Virol*, **69**, 2101 - 2109.
76. Damgaard, C.K., Dyhr-Mikkelsen, H. and Kjems, J. (1998) Mapping the RNA binding sites for human immunodeficiency virus type-1 gag and NC proteins within the complete HIV-1 and -2 untranslated leader regions. *Nucleic Acids Res*, **26**, 3667-3676.
77. Rist, M.J. and Marino, J.P. (2002) Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1. *Biochemistry*, **41**, 14762-14770.
78. Mirambeau, G., Lyonnais, S., Coulaud, D., Hameau, L., Lafosse, S., Jeusset, J., Justome, A., Delain, E., Gorelick, R.J. and Le Cam, E. (2006) Transmission electron microscopy reveals an optimal HIV-1 nucleocapsid aggregation with single-stranded nucleic acids and the mature HIV-1 nucleocapsid protein. *J Mol Biol*, **364**, 496-511.
79. Paillart, J.-C., Westhof, E., Ehresmann, C., Ehresmann, B. and Marquet, R. (1997) Non-canonical interactions in a kissing loop complex: the dimerization initiation site of HIV-1 genomic RNA. *Journal of Molecular Biology*, **270**, 36-49.
80. Dardel, F., Marquet, R., Ehresmann, C., Ehresmann, B. and Blanquet, S. (1998) Solution studies of the dimerization initiation site of HIV-1 genomic RNA. *Nucleic Acids Res*, **26**, 3567-3571.
81. Mujeeb, A., Clever, J.L., Billeci, T.M., James, T.L. and Parslow, T.G. (1998) Structure of the dimer initiation complex of HIV-1 genomic RNA. *Nat Struct Biol*, **5**, 432-436.
82. Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C. and Dumas, P. (2001) Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat Struct Biol*, **8**, 1064-1068.
83. Ennifar, E. and Dumas, P. (2006) Polymorphism of bulged-out residues in HIV-1 RNA DIS kissing complex and structure comparison with solution studies. *J Mol Biol*, **356**, 771-782.
84. Kieken, F., Paquet, F., Brule, F., Paoletti, J. and Lancelot, G. (2006) A new NMR solution structure of the SL1 HIV-1Lai loop-loop dimer. *Nucleic Acids Res*, **34**, 343-352.
85. Baba, S., Takahashi, K., Noguchi, S., Takaku, H., Koyanagi, Y., Yamamoto, N. and Kawai, G. (2005) Solution RNA structures of the HIV-1 dimerization initiation site in the kissing-loop and extended-duplex dimers. *J Biochem*, **138**, 583-592.

86. Ennifar, E., Yusupov, M., Walter, P., Marquet, R., Ehresmann, B., Ehresmann, C. and Dumas, P. (1999) The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. *Structure*, **7**, 1439-1449.
87. Girard, F., Barbault, F., Gouyette, C., Huynh-Dinh, T., Paoletti, J. and Lancelot, G. (1999) Dimer initiation sequence of HIV-1Lai genomic RNA: NMR solution structure of the extended duplex. *J Biomol Struct Dyn*, **16**, 1145-1157.
88. Mujeeb, A., Parslow, T.G., Zarrinpar, A., Das, C. and James, T.L. (1999) NMR structure of the mature dimer initiation complex of HIV-1 genomic RNA. *FEBS Lett*, **458**, 387-392.
89. Ulyanov, N.B., Mujeeb, A., Du, Z., Tonelli, M., Parslow, T.G. and James, T.L. (2006) NMR structure of the full-length linear dimer of stem-loop-1 RNA in the HIV-1 dimer initiation site. *J Biol Chem*, **281**, 16168-16177.
90. Aci, S., Mazier, S. and Genest, D. (2005) Conformational pathway for the kissing complex-->extended dimer transition of the SL1 stem-loop from genomic HIV-1 RNA as monitored by targeted molecular dynamics techniques. *J Mol Biol*, **351**, 520-530.
91. Aci, S., Gangneux, L., Paoletti, J. and Genest, D. (2004) On the stability of different experimental dimeric structures of the SL1 sequence from the genomic RNA of HIV-1 in solution: a molecular dynamics simulation and electrophoresis study. *Biopolymers*, **74**, 177-188.
92. Beaurain, F. and Laguerre, M. (2003) MD studies of the DIS/DIS kissing complex solution and x-ray structures. *Oligonucleotides*, **13**, 501-514.
93. Reblova, K., Fadrna, E., Sarzynska, J., Kulinski, T., Kulhanek, P., Ennifar, E., Koca, J. and Sponer, J. (2007) Conformations of flanking bases in HIV-1 RNA DIS kissing complexes studied by molecular dynamics. *Biophys J*, **93**, 3932-3949.
94. Sarzynska, J., Reblova, K., Sponer, J. and Kulinski, T. (2008) Conformational transitions of flanking purines in HIV-1 RNA dimerization initiation site kissing complexes studied by CHARMM explicit solvent molecular dynamics. *Biopolymers*, **89**, 732-746.
95. Greatorex, J., Gallego, J., Varani, G. and Lever, A. (2002) Structure and stability of wild-type and mutant RNA internal loops from the SL-1 domain of the HIV-1 packaging signal. *J Mol Biol*, **322**, 543 - 557.
96. Mujeeb, A., Ulyanov, N.B., Georgantis, S., Smirnov, I., Chung, J., Parslow, T.G. and James, T.L. (2007) Nucleocapsid protein-mediated maturation of dimer initiation complex of full-length SL1 stemloop of HIV-1: sequence effects and mechanism of RNA refolding. *Nucleic Acids Res*, **35**, 2026-2034.
97. Mazier, S. and Genest, D. (2008) Insight into the intrinsic flexibility of the SL1 stem-loop from genomic RNA of HIV-1 as probed by molecular dynamics simulation. *Biopolymers*, **89**, 187-196.
98. Lawrence, D.C., Stover, C.C., Noznitsky, J., Wu, Z. and Summers, M.F. (2003) Structure of the intact stem and bulge of HIV-1 Psi-RNA stem-loop SL1. *J Mol Biol*, **326**, 529-542.
99. Yuan, Y., Kerwood, D.J., Paoletti, A.C., Shubsda, M.F. and Borer, P.N. (2003) Stem of SL1 RNA in HIV-1: structure and nucleocapsid protein binding for a 1 x 3 internal loop. *Biochemistry*, **42**, 5259-5269.

100. Mazier, S. and Genest, D. (2007) Molecular dynamics simulation for probing the flexibility of the 35 nucleotide SL1 sequence kissing complex from HIV-1Lai genomic RNA. *J Biomol Struct Dyn*, **24**, 471-479.

101. Darlix, J.L., Lapadat-Tapolsky, M., de Rocquigny, H. and Roques, B.P. (1995) First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *J Mol Biol*, **254**, 523-537.

102. Rein, A., Henderson, L.E. and Levin, J.G. (1998) Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem Sci*, **23**, 297 - 301.

103. Urbaneja, M.A., Kane, B.P., Johnson, D.G., Gorelick, R.J., Henderson, L.E. and Casas-Finet, J.R. (1999) Binding properties of the human immunodeficiency virus type 1 nucleocapsid protein p7 to a model RNA: elucidation of the structural determinants for function. *J Mol Biol*, **287**, 59-75.

104. Berkowitz, R., Fisher, J. and Goff, S.P. (1996) RNA packaging. *Curr Top Microbiol Immunol*, **214**, 177 - 218.

105. Karpel, R.L., Henderson, L.E. and Oroszlan, S. (1987) Interactions of retroviral structural proteins with single-stranded nucleic acids. *J Biol Chem*, **262**, 4961-4967.

106. Dib-Hajj, F., Khan, R. and Giedroc, D.P. (1993) Retroviral nucleocapsid proteins possess potent nucleic acid strand renaturation activity. *Protein Sci*, **2**, 231-243.

107. Guo, J., Henderson, L.E., Bess, J., Kane, B. and Levin, J.G. (1997) Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA. *J Virol*, **71**, 5178-5188.

108. You, J.C. and McHenry, C.S. (1994) Human immunodeficiency virus nucleocapsid protein accelerates strand transfer of the terminally redundant sequences involved in reverse transcription. *J Biol Chem*, **269**, 31491-31495.

109. Feng, Y.X., Copeland, T.D., Henderson, L.E., Gorelick, R.J., Bosche, W.J., Levin, J.G. and Rein, A. (1996) HIV-1 nucleocapsid protein induces "maturation" of dimeric retroviral RNA in vitro. *Proc Natl Acad Sci U S A*, **93**, 7577 - 7581.

110. Tanchou, V., Delaunay, T., Bodeus, M., Roques, B., Darlix, J.L. and Benarous, R. (1995) Conformational changes between human immunodeficiency virus type 1 nucleocapsid protein NCp7 and its precursor NCp15 as detected by anti-NCp7 monoclonal antibodies. *J Gen Virol*, **76 (Pt 10)**, 2457-2466.

111. Lapadat-Tapolsky, M., De Rocquigny, H., Van Gent, D., Roques, B., Plasterk, R. and Darlix, J.L. (1993) Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle. *Nucleic Acids Res*, **21**, 831-839.

112. Huang, Y., Khorchid, A., Wang, J., Parniak, M.A., Darlix, J.L., Wainberg, M.A. and Kleiman, L. (1997) Effect of mutations in the nucleocapsid protein (NCp7) upon Pr160(gag-pol) and tRNA(Lys) incorporation into human immunodeficiency virus type 1. *J Virol*, **71**, 4378-4384.

113. Cameron, C.E., Ghosh, M., Le Grice, S.F.J. and Benkovic, S.J. (1997) Mutations in HIV reverse transcriptase which alter RNase H activity and decrease strand transfer efficiency are suppressed by HIV nucleocapsid 鈥塸 rotein. *Proceedings*

of the National Academy of Sciences of the United States of America, **94**, 6700-6705.

114. Carteau, S., Batson, S.C., Poljak, L., Mouscadet, J.F., de Rocquigny, H., Darlix, J.L., Roques, B.P., Kas, E. and Auclair, C. (1997) Human immunodeficiency virus type 1 nucleocapsid protein specifically stimulates Mg2+-dependent DNA integration in vitro. *J Virol*, **71**, 6225-6229.
115. Coffin, J., Haase, A., Levy, J.A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H., Vogt, P. *et al.* (1986) Human immunodeficiency viruses. *Science*, **232**, 697.
116. Coffin, J., Haase, A., Levy, J.A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H., Vogt, P. *et al.* (1986) What to call the AIDS virus? *Nature*, **321**, 10.
117. Khan, R. and Giedroc, D.P. (1992) Recombinant human immunodeficiency virus type 1 nucleocapsid (NCp7) protein unwinds tRNA. *J Biol Chem*, **267**, 6689-6695.
118. Morellet, N., de Rocquigny, H., Mely, Y., Jullian, N., Demene, H., Ottmann, M., Gerard, D., Darlix, J.L., Fournie-Zaluski, M.C. and Roques, B.P. (1994) Conformational behaviour of the active and inactive forms of the nucleocapsid NCp7 of HIV-1 studied by 1H NMR. *J Mol Biol*, **235**, 287-301.
119. Summers, M.F., Henderson, L.E., Chance, M.R., Bess, J.W., Jr., South, T.L., Blake, P.R., Sagi, I., Perez-Alvarado, G., Sowder, R.C., 3rd, Hare, D.R. *et al.* (1992) Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1. *Protein Sci*, **1**, 563-574.
120. Dorfman, T., Luban, J., Goff, S.P., Haseltine, W.A. and Gottlinger, H.G. (1993) Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein. *J Virol*, **67**, 6159 - 6169.
121. Luban, J. and Goff, S.P. (1994) Mutational analysis of cis-acting packaging signals in human immunodeficiency virus type 1 RNA. *J Virol*, **68**, 3784 - 3793.
122. Berkowitz, R.D., Luban, J. and Goff, S.P. (1993) Specific binding of human immunodeficiency virus type 1 gag polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays. *J Virol*, **67**, 7190 - 7200.
123. Berkowitz, R.D., Hammarskjold, M.L., Helga-Maria, C., Rekosh, D. and Goff, S.P. (1995) 5' regions of HIV-1 RNAs are not sufficient for encapsidation: implications for the HIV-1 packaging signal. *Virology*, **212**, 718 - 723.
124. Song, R., Kafaie, J., Yang, L. and Laughrea, M. (2007) HIV-1 viral RNA is selected in the form of monomers that dimerize in a three-step protease-dependent process; the DIS of stem-loop 1 initiates viral RNA dimerization. *J Mol Biol*, **371**, 1084-1098.
125. Kafaie, J., Song, R., Abrahamyan, L., Mouland, A.J. and Laughrea, M. (2008) Mapping of nucleocapsid residues important for HIV-1 genomic RNA dimerization and packaging. *Virology*, **375**, 592-610.
126. Theilleux-Delalande, V., Girard, F., Huynh-Dinh, T., Lancelot, G. and Paoletti, J. (2000) The HIV-1(Lai) RNA dimerization. Thermodynamic parameters

associated with the transition from the kissing complex to the extended dimer. *Eur J Biochem*, **267**, 2711-2719.

127. Mihailescu, M.R. and Marino, J.P. (2004) A proton-coupled dynamic conformational switch in the HIV-1 dimerization initiation site kissing complex. *Proc Natl Acad Sci U S A*, **101**, 1189-1194.
128. Lodmell, J.S., Ehresmann, C., Ehresmann, B. and Marquet, R. (2000) Convergence of natural and artificial evolution on an RNA loop-loop interaction: the HIV-1 dimerization initiation site. *Rna*, **6**, 1267-1276.
129. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley, New York.
130. Akke, M. (2002) NMR methods for characterizing microsecond to millisecond dynamics in recognition and catalysis. *Curr Opin Struct Biol*, **12**, 642-647.
131. Kern, D. and Zuiderweg, E.R. (2003) The role of dynamics in allosteric regulation. *Curr Opin Struct Biol*, **13**, 748-757.
132. Palmer, A.G. (1997) Probing molecular motion by NMR. *Current Opinion in Structural Biology*, **7**, 732-737.
133. Palmer, A.G., 3rd. (2004) NMR characterization of the dynamics of biomacromolecules. *Chem Rev*, **104**, 3623-3640.
134. Palmer, A.G., Williams, J. and McDermott, A. (1996) Nuclear magnetic resonance studies of biopolymer dynamics. *Journal of Physical Chemistry*, **100**, 13293-13310.
135. Kay, L.E. (1998) Protein dynamics from NMR. *Nature Structural Biology*, **5**, 513-517.
136. Al-Hashimi, H.M. (2005) Dynamics-Based Amplification of RNA Function and Its Characterization by Using NMR Spectroscopy. *Chembiochem*, **6**, 1506-1519.
137. Mittermaier, A. and Kay, L.E. (2006) New tools provide new insights in NMR studies of protein dynamics. *Science*, **312**, 224-228.
138. Henzler-Wildman, K. and Kern, D. (2007) Dynamic personalities of proteins. *Nature*, **450**, 964-972.
139. Farrow, N.A., Zhang, O., Forman-Kay, J.D. and Kay, L.E. (1994) A Heteronuclear Correlation Experiment for Simultaneous Determination of 15N Longitudinal Decay and Chemical Exchange Rates of Systems in Slow Equilibrium. *Journal of Biomolecular NMR*, **4**, 727-734.
140. Akke, M. and III, A.G.P. (1996) Monitoring macromolecular motions on microsecond to millisecond time scales by R(1)rho-R(1) constant relaxation time NMR spectroscopy. *J. Am. Chem. Soc.*, **118**, 911-912.
141. Bruschweiler, R. (2003) New approaches to the dynamic interpretation and prediction of NMR relaxation data from proteins. *Curr Opin Struct Biol*, **13**, 175-183.
142. Al-Hashimi, H.M. and Patel, D.J. (2002) Residual dipolar couplings: Synergy between NMR and structural genomics. *Journal of Biomolecular Nmr*, **22**, 1-8.
143. Al-Hashimi, H.M., Gosser, Y., Gorin, A., Hu, W., Majumdar, A. and Patel, D.J. (2002) Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. *J Mol Biol*, **315**, 95-102.

144. Getz, M., Sun, X., Casiano-Negroni, A., Zhang, Q. and Al-Hashimi, H.M. (2007) NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings. *Biopolymers*, **86**, 384-402.
145. Tolman, J.R., Al-Hashimi, H.M., Kay, L.E. and Prestegard, J.H. (2001) Structural and dynamic analysis of residual dipolar coupling data for proteins. *Journal of the American Chemical Society*, **123**, 1416-1424.
146. Tolman, J.R., Flanagan, J.M., Kennedy, M.A. and Prestegard, J.H. (1997) NMR evidence for slow collective motions in cyanometmyoglobin. *Nature Structural Biology*, **4**, 292-297.
147. Peti, W., Meiler, J., Bruschweiler, R. and Griesinger, C. (2002) Model-free analysis of protein backbone motion from residual dipolar couplings. *J Am Chem Soc*, **124**, 5822-5833.
148. Furtig, B., Buck, J., Manoharan, V., Bermel, W., Jaschke, A., Wenter, P., Pitsch, S. and Schwalbe, H. (2007) Time-resolved NMR studies of RNA folding. *Biopolymers*, **86**, 360-383.
149. Furtig, B., Richter, C., Wohnert, J. and Schwalbe, H. (2003) NMR spectroscopy of RNA. *Chembiochem*, **4**, 936-962.
150. Kellogg, G.W. and Schweitzer, B.I. (1993) Two- and three-dimensional 31P-driven NMR procedures for complete assignment of backbone resonances in oligodeoxyribonucleotides. *J Biomol NMR*, **3**, 577-595.
151. Sklenar, V., Dieckmann, T., Butcher, S.E. and Feigon, J. (1996) Through-bond correlation of imino and aromatic resonances in 13C-, 15N-labeled RNA via heteronuclear TOCSY. *J Biomol NMR*, **7**, 83-87.
152. Sklenar, V., Peterson, R.D., Rejante, M.R. and Feigon, J. (1993) Two- and three-dimensional HCN experiments for correlating base and sugar resonances in 15N,13C-labeled RNA oligonucleotides. *J Biomol NMR*, **3**, 721-727.
153. Lipari, G. and Szabo, A. (1981) Nuclear magnetic resonance relaxation in nucleic acid fragments: models for internal motion. *Biochemistry*, **20**, 6250-6256.
154. Kay, L.E., Torchia, D.A. and Bax, A. (1989) Backbone Dynamics of Proteins As Studied by ^{15}N Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphlococcal Nuclease. *Biochemistry*, **28**, 8972-8979.
155. Akke, M., Fiala, R., Jiang, F., Patel, D. and Palmer, A.G. (1997) Base dynamics in a UUCG tetraloop RNA hairpin characterized by N-15 spin relaxation: Correlations with structure and stability. *RNA*, **3**, 702-709.
156. Abragam, A. (1961) *Principles of Nuclear Magnetism*. Clarendon Press, Oxford.
157. Lipari, G. and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *Journal of the American Chemical Society*, **104**, 4546-4559.
158. Clore, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C. and Gronenborn, A.M. (1990) Deviations from the Simple Two-Parameter Model-Free Approach to the Interpretation of Nitrogen-15 Nuclear Magnetic Relaxation of Proteins. *Journal of the American Chemical Society*, **112**, 4989-4991.
159. Yip, G.N. and Zuiderweg, E.R. (2004) A phase cycle scheme that significantly suppresses offset-dependent artifacts in the R2-CPMG 15N relaxation experiment. *J Magn Reson*, **171**, 25-36.

160. Grzesiek, S. and Bax, A. (1993) The Importance of Not Saturating H_2O in Protein NMR. Applications to Sensitivity Enhancement and NOE Measurements. *J. Am. Chem. Soc.*, **115**, 12593-12594.
161. Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D. and Kay, L.E. (1994) Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by 15N NMR Relaxation. *Biochemistry*, **33**, 5984-6003.
162. Tjandra, N. and Bax, A. (1997) Measurement of dipolar contributions to (1)J(CH) splittings from magnetic-field dependence of J modulation in two- dimensional NMR spectra. *Journal of Magnetic Resonance*, **124**, 512-515.
163. Bailor, M.H., Musselman, C., Hansen, A.L., Gulati, K., Patel, D.J. and Al-Hashimi, H.M. (2007) Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings. *Nat Protoc*, **2**, 1536-1546.
164. Prestegard, J.H., Tolman, J.R., Al-Hashimi, H.M. and Andrec, M. (1999) In Krishna, N. R. and Berliner, L. J. (eds.), *Biological Magnetic Resonance*. Plenum, New York, Vol. 17, pp. 311-355.
165. Prestegard, J.H., Al-Hashimi, H.M. and Tolman, J.R. (2000) NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Quarterly Reviews of Biophysics*, **33**, 371-424.
166. Musselman, C., Pitt, S.W., Gulati, K., Foster, L.L., Andricioaei, I. and Al-Hashimi, H.M. (2006) Impact of static and dynamic A-form heterogeneity on the determination of RNA global structural dynamics using NMR residual dipolar couplings. *J Biomol NMR*, **36**, 235-249.
167. Zhang, Q., Sun, X., Watt, E.D. and Al-Hashimi, H.M. (2006) Resolving the motional modes that code for RNA adaptation. *Science*, **311**, 653-656.
168. Zhang, Q., Stelzer, A.C., Fisher, C.K. and Al-Hashimi, H.M. (2007) Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature*, **450**, 1263-1267.
169. Scott, W.G. (2001) Ribozyme catalysis via orbital steering. *J Mol Biol*, **311**, 989-999.
170. Scott, W.G. (1999) RNA structure, metal ions, and catalysis. *Curr Opin Chem Biol*, **3**, 705-709.
171. Scott, W.G., Finch, J.T. and Klug, A. (1995) The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell*, **81**, 991-1002.
172. Scott, W.G., Murray, J.B., Arnold, J.R., Stoddard, B.L. and Klug, A. (1996) Capturing the structure of a catalytic RNA intermediate: the hammerhead ribozyme. *Science*, **274**, 2065-2069.
173. Misra, V.K. and Draper, D.E. (2002) The linkage between magnesium binding and RNA folding. *J Mol Biol*, **317**, 507-521.
174. Misra, V.K. and Draper, D.E. (2000) Mg(2+) binding to tRNA revisited: the nonlinear Poisson-Boltzmann model. *J Mol Biol*, **299**, 813-825.
175. Misra, V.K. and Draper, D.E. (2001) A thermodynamic framework for Mg2+ binding to RNA. *Proc Natl Acad Sci U S A*, **98**, 12456-12461.
176. Draper, D.E. (2004) A guide to ions and RNA structure. *RNA*, **10**, 335-343.

177. Garcia-Garcia, C. and Draper, D.E. (2003) Electrostatic interactions in a peptide--RNA complex. *J Mol Biol*, **331**, 75-88.
178. Draper, D.E., Grilley, D. and Soto, A.M. (2005) Ions and RNA folding. *Annu Rev Biophys Biomol Struct*, **34**, 221-243.
179. Hansen, M.R., Simorre, J.P., Hanson, P., Mokler, V., Bellon, L., Beigelman, L. and Pardi, A. (1999) Identification and characterization of a novel high affinity metal-binding site in the hammerhead ribozyme. *Rna*, **5**, 1099-1104.
180. Butcher, S.E., Allain, F.H. and Feigon, J. (2000) Determination of metal ion binding sites within the hairpin ribozyme domains by NMR. *Biochemistry*, **39**, 2174-2182.
181. Ott, G., Arnold, L. and Limmer, S. (1993) Proton NMR studies of manganese ion binding to tRNA-derived acceptor arm duplexes. *Nucleic Acids Res*, **21**, 5859-5864.
182. Bertini, I., Gupta, Y.K., Luchinat, C., Parigi, G., Schlorb, C. and Schwalbe, H. (2005) NMR spectroscopic detection of protein protons and longitudinal relaxation rates between 0.01 and 50 MHz. *Angew Chem Int Ed Engl*, **44**, 2223-2225.
183. Buck, J., Furtig, B., Noeske, J., Wohnert, J. and Schwalbe, H. (2007) Time-resolved NMR methods resolving ligand-induced RNA folding at atomic resolution. *Proc Natl Acad Sci U S A*, **104**, 15699-15704.
184. Howe, P.W., Nagai, K., Neuhaus, D. and Varani, G. (1994) NMR studies of U1 snRNA recognition by the N-terminal RNP domain of the human U1A protein. *Embo J*, **13**, 3873-3881.
185. Fourmy, D., Guittet, E. and Yoshizawa, S. (2002) Structure of Prokaryotic SECIS mRNA Hairpin and its Interaction with Elongation Factor SelB. *Journal of Molecular Biology*, **324**, 137-150.
186. Lane, A.N., Kelly, G., Ramos, A. and Frenkiel, T.A. (2001) Determining binding sites in protein–nucleic acid complexes by cross-saturation. *Journal of Biomolecular NMR*, **21**, 127-139.

Chapter 2

Characterizing mutant monomer SL1m RNA structural dynamics in the absence and presence of Mg^{2+}

2.1 Introduction

Despite the involvement of SL1 in numerous functional structural transitions and ample evidence indicating that the inherent flexibility in SL1 is important in driving these conformational changes, little is known about the dynamical properties of SL1 in kissing or duplex form. Furthermore, the far majority of studies have focused on SL1 constructs that lack the internal loop, which is believed to give rise to functionally important dynamics. Two NMR studies of monomeric SL1 constructs containing the internal loop provide conflicting views of the internal motions (1,2). Finally, the metal binding properties of the SL1 internal loop have not been explored and their potential role in isomerization remains to be elucidated.

In this study we examine an SL1 monomer that contains the internal loop via a combination of NMR techniques, including residual dipolar couplings (RDCs) for probing internal motions at sub-millisecond timescales, dynamically decoupled spin relaxation to probe internal motions at nanosecond and faster timescales, and chemical shift mapping to quantitatively characterize the internal flexibility of the internal internal loop and how it varies upon Mg^{2+} binding.

2.2 Results

2.2.1 Design monomeric SL1m construct as a model for the stem-internal loop-stem element of the kissing SL1 dimer

We characterized the stem-internal loop-stem element of SL1 using a previously reported SL1m (2) that is impaired from forming dimers through replacement of the wild-type apical loop with a GAGA tetraloop (Figure 2-1A). The lower stem I is also extended by three-pairs to optimize synthesis by *in vitro* transcription (2). A previous NMR study used spectral comparisons to show that replacement of the wild-type apical loop with a UUCG tetraloop does not impact the structural integrity of the stem-internal loop-stem element of the SL1 kissing dimer (3). We examined if this was also the case for the GAGA-loop SL1m mutant by comparing its spectra with those of the corresponding kissing SL1 dimer (Figure 2-1B and C). Aside from the expected differences at residues near the loop (Figure 2-1B and C, highlighted in red), very good chemical shift agreement was observed, indicating that SL1m recapitulates the stem-internal loop-stem element in the kissing SL1 dimer (2). As we discuss in chapter 3, similar exchange broadening and Mg^{2+}-induced chemical shift perturbations were also observed in the two constructs, indicating that SL1m also retains the dynamical and Mg^{2+} binding properties of the stem-internal loop-stem element in the SL1 kissing dimer.

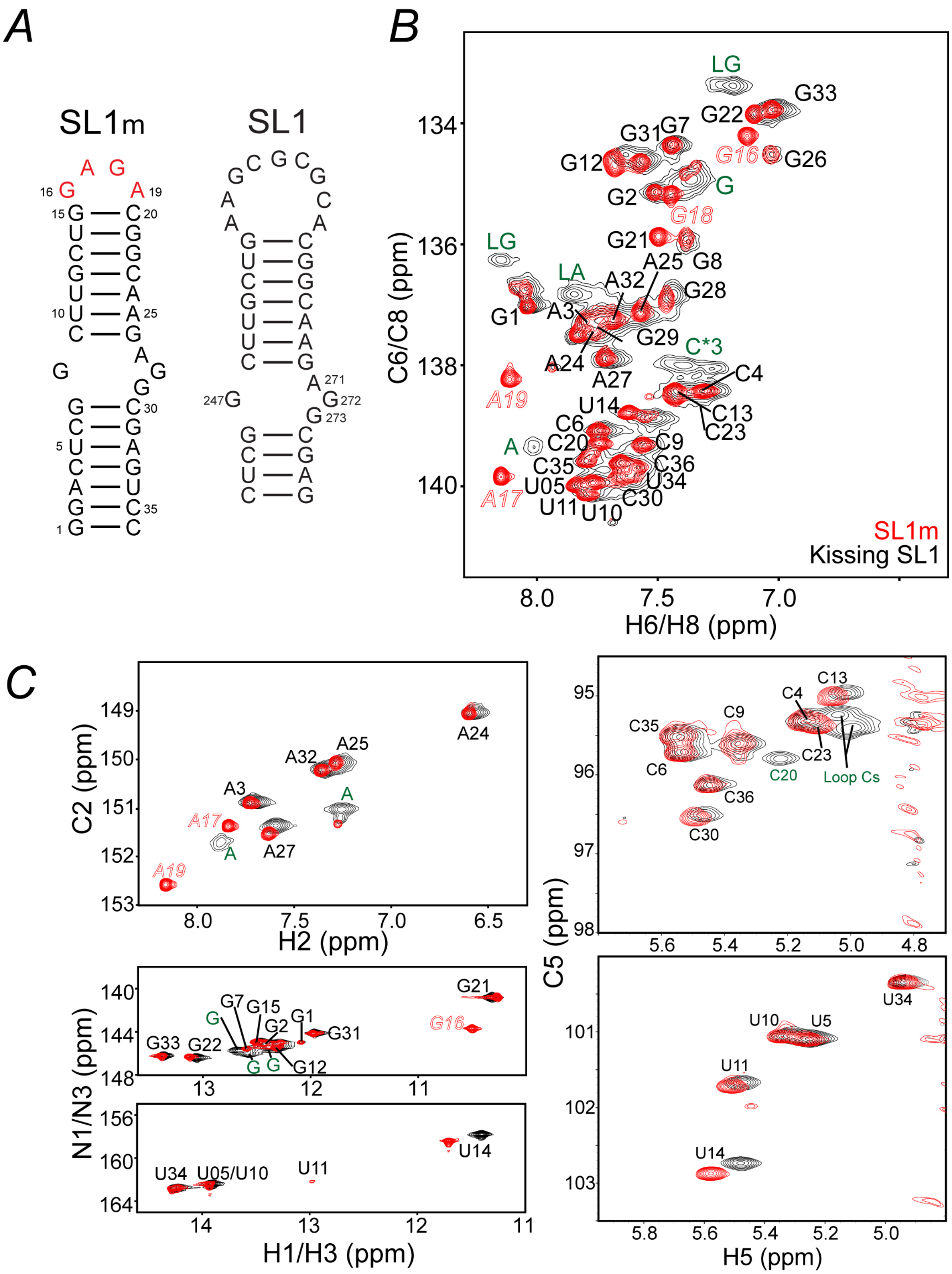

Figure 2-1. Comparison of spectra for the SL1m monomer and kissing SL1 dimer. (A) Secondary structures of SL1m and SL1. (B-C) Overlay of 2D ^{1}H-^{13}C of and ^{1}H-^{15}N HSQC spectra of SL1m (in red) and SL1 kissing dimer (in black) in the absence of Mg^{2+}.

2.2.2 SL1 dynamically interconverts between alternative secondary structures causing partial melting of base-pairs in the upper stem in the absence of Mg^{2+}

In a previous study of a UUCG-loop SL1 monomer, internal loop resonances exhibited severe exchange broadening characteristic of micro-to-millisecond timescale motions and this precluded detailed NMR structural characterization (1). Replacement of the wild-type G-AGG internal loop with a G-GGA internal loop yielded an SL1 construct with far less broadening and higher (~8°C) UV melting temperatures (1). In contrast, exchange broadening was not observed in the Summers et. al's study of the GAGA-loop SL1m (2). Given that similar NMR buffer conditions were used in the two studies (absence of divalent ions), the source of this discrepancy remains unresolved.

In our study, we observed exchange broadening in resonances belonging to the internal loop and neighboring residues in the absence of Mg^{2+} (Figure 2-2A and B). The exchange broadening was also observed in the SL1 kissing dimer (Figure 2-1B and C) and in SL1m constructs in which stem I was elongated by twenty-two base-pairs (see Figure 2-7A). The broadening was also observed when using the same buffer (10mM Tris-d11, pH ~ 8.0, 0.1 mM EDTA in $^{2}H_{2}O$) used in the previous NMR study of GAGA-loop SL1m (data not shown) (2). The broadening was however strongly dependent on temperature (Figure 2-2A). It was very pronounced at 5°C in which the dynamic process appears to be in intermediate/slow exchange and was virtually undetectable at 45°C in which the process appears to be in fast exchange (Figure 2-2B). In this regard, it is possible that exchange broadening was not observed in the previous study of GAGA-loop SL1m because a relatively high temperature of 37°C was used to record NMR data (2).

Residues that experience exchange broadening in SL1m likely sample more than one conformation. In addition to all internal loop residues, broadening was observed at C30 below the internal loop in stem I and the three base-pairs above the internal loop in stem II (C9-G26, U10-A25 and U11-A24) (Figure 2-2B). These base-pairs, including all three base-pairs above the internal loop in stem II, also have non-observable or weak imino signals indicating that their hydrogen alignments deviates from ideality (Figure 2-1). The observed broadening and hydrogen bond melting pattern is consistent with a dynamical equilibrium involving two distinct SL1m secondary structures (referred to as "A" and "B", Figure 2-2C) that was previously proposed based on secondary structure prediction (3). Conformer A contains an asymmetric internal loop and is observed in all NMR structures reported to date (Figure 2-2C). Conformer B contains an internal loop and symmetric internal loop and has not been directly observed in any structural studies. The A to B transition, possibly via one or more intermediates (I), involves replacing Watson-Crick residues A25 and G26 in conformer A by internal loop residues A27 and G28 (Figure 2-2C). This leads to changes in the local environment of precisely those residues for which significant exchange broadening is observed, namely the internal loop and three base-pairs above it in stem II (Figure 2-2C, shown in red). It also leads to transient melting/reannealing of the two upper base-pairs (C9-G26 and U10-A25) explaining why their imino signals are exchange broadened beyond the limit of detection (Figure 2-2C). In this context, the reduced exchange broadening observed in the G-GGA SL1 mutant (1) also can be attributed in part to its inability to slip into the B conformer. Together, these results suggest that the SL1m internal loop transiently unzips the upper stem by stabilizing more than one conformation involving distinct hydrogen alignments.

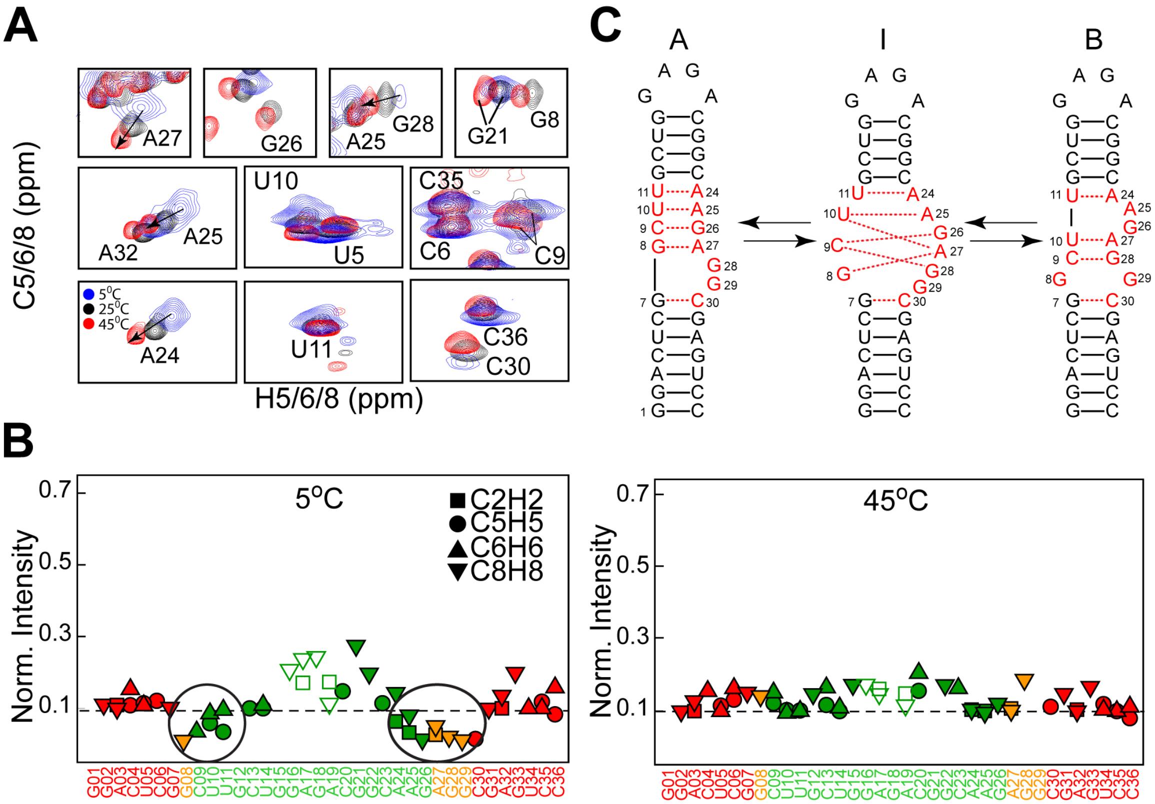

Figure 2-2. Conformational exchange in SL1m. (A) Representative spectra showing the temperature dependence of the exchange broadening at 5°C (in blue), 25°C (in black), and 45°C (in red). (B) Relative resonance intensities of non-exchangeable C-H resonances measured in SL1m at 5°C and 45°C. The intensities for a given type of bond vector are normalized to a minimum value of 0.1 as previously described (4). (C) Proposed dynamical equilibrium between SL1 conformers denoted “A” and “B” and possible intermediates denoted “I” along the pathway involving a smaller number of hydrogen bond alignments. Residues undergoing the largest degree of exchange broadening are shown in red.

2.2.3 Mg^{2+} binds the internal loop, arrests the dynamical exchange, and stabilizes the upper stem

All NMR structures reported thus far for the internal loop-containing SL1 have been determined in the absence of divalent ions. To examine if Mg^{2+} binds to the SL1m internal loop, we performed chemical shift titration experiments in which 2D HSQC spectra were recorded following incremental addition of Mg^{2+} up to 5 mM. Very large chemical shift perturbations were observed that are indicative of specific Mg^{2+} binding

(Figure 2-3B). The apparent K_ds obtained from fitting the chemical shift data to a two state binding model cluster within a relatively narrow range (0.9 - 1.9 mM) indicative of tight Mg^{2+} binding (Figure 2-3C).

Interestingly, the largest chemical shift perturbations, which reflect Mg^{2+} localization effects and/or Mg^{2+} induced changes in SL1 conformation, were observed precisely at the residues that exhibit marked exchange broadening in the internal loop and the upper stem (Figure 2-3B). The exchange broadening at these sites is markedly reduced upon Mg^{2+} binding (Figure 2-3A). This is accompanied by the appearance and intensification of imino signals belonging to U10 and U11, respectively, in the upper stem (Figure 2-3A) for which an A-U Watson-Crick hydrogen bond alignment could be established directly using the J_{NN}-COSY experiment (5,6). An additional Watson-Crick guanine imino signal appears which could not be directly assigned but which may correspond to G26 (lack of NOE connectives with U10 is not surprising given that the U10 signal remains somewhat weak). Similar spectral changes were observed for the kissing SL1 dimer (Figure 2-1, also see chapter 3). These results suggest that Mg^{2+} binds and stabilizes the SL1m internal loop and in doing so arrests the dynamical exchange and partial melting of the upper stem.

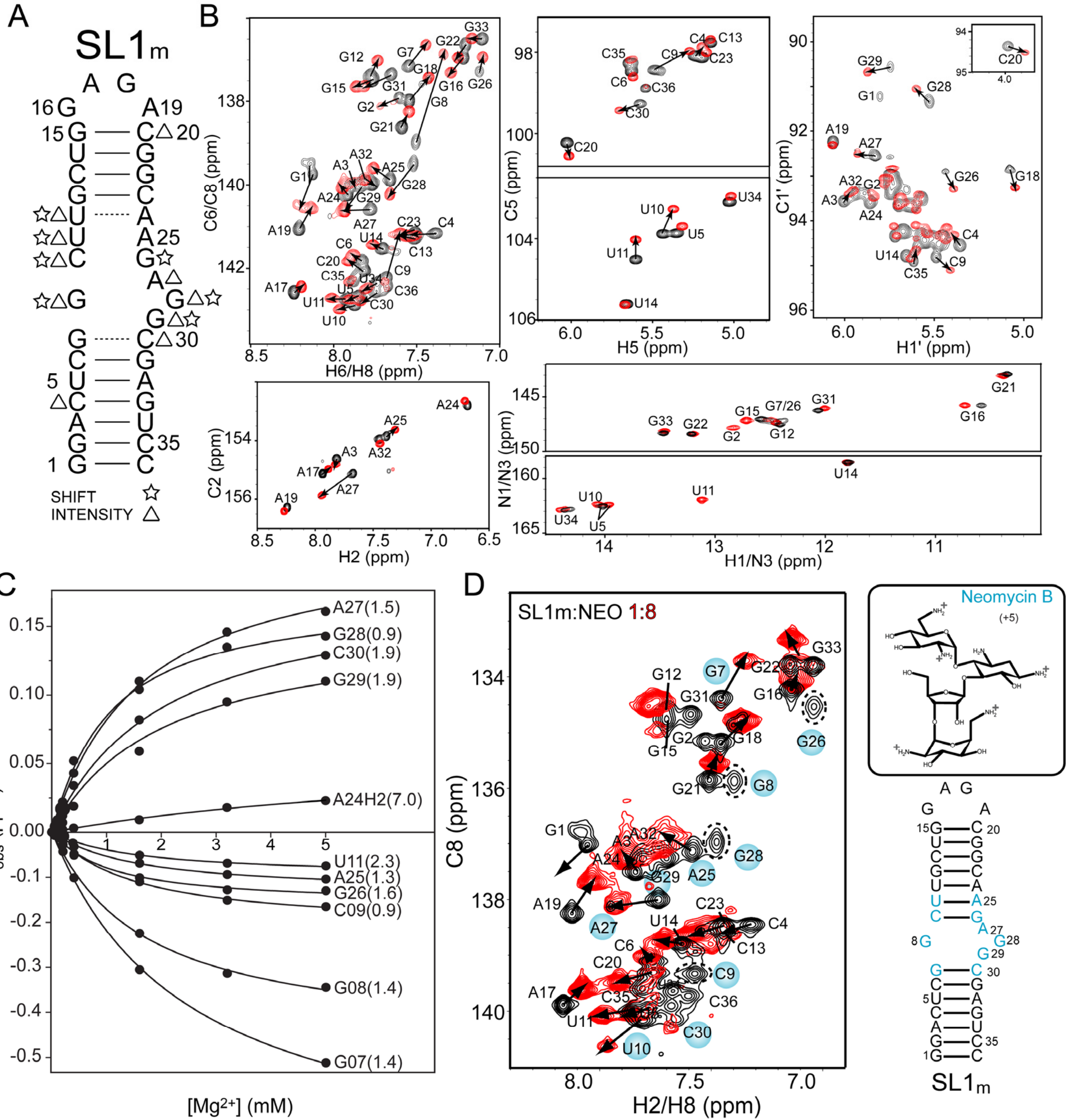

Figure 2-3. Chemical shift mapping of Mg^{2+} binding to the SL1m internal loop. (A) Residues undergoing the largest Mg^{2+} induced chemical shift perturbations (top 20% for a given type of resonance) and that yield the tightest binding (K_d < 2.0 mM) are highlighted on the SL1m secondary structure in red and and black box, respectively. (B) Overlay of 2D ^1H-^{13}C and ^1H-^{15}N HSQC spectra of SL1m recorded in the absence (in black) and presence (in red) of 5 mM Mg^{2+}. (C) Representative titration curves as a function of total Mg^{2+} concentration with K_d values shown at the end of each curve. (D) Neomycin B molecule and overlay of 2D ^1H-^{13}C HSQC spectra of SL1m recorded in the absence (in black) and presence (in red) of 2.4 mM neomycin. Residues undergoing the largest neomycin B induced chemical shift perturbations are highlighted in light blue.

2.2.4 Local conformation of the stems in the absence and presence of Mg^{2+}

We measured RDCs in the Pf1 phage ordering medium (7,8) and used them to characterize the local conformation of the stems in the absence and presence of Mg^{2+} (Table 2-1). In particular, we used the order tensor analysis (9-12) to examine the fit between measured RDCs and an idealized A-form helical structure for the two stems. This provided a basis for evaluating whether conformer A or B is the dominant species in solution (Figure 2-2C). RDC studies (10,13-19) together with a recent survey of high-resolution X-ray structures of RNA (10) show the idealized A-form helix to be an excellent model geometry for non-terminal Watson-Crick base-pairs in different RNA contexts.

Initially, only RDCs belonging to hydrogen bonded Watson-Crick base-pairs as inferred from the imino 2D ^{15}N-^{1}H HSQC spectra were included in the order tensor fit which was carried out independently for stems I and II. The best-fit order tensor and idealized A-form geometry were then used to back-predict all of the measured RDCs including ones not included in the fit. For both stems, excellent agreement was obtained for RDCs measured in the hydrogen-bonded base-pairs both in the absence and presence of Mg^{2+} (Figure 2-4, filled symbols). The observed root-mean-square-deviation (RMSD) between measured and predicted RDCs (0.6-1.4 Hz) is well within the estimated RDC experimental uncertainty (~ 2.5 Hz). In contrast, the agreement was variable for the non-hydrogen bonded base-pairs that were excluded from the order tensor fit (Figure 2-4, open symbols).

For stem I, very poor agreement was observed for terminal residue C30 immediately below the internal loop in the absence (Figure 2-4A) but not in the presence of Mg^{2+} (Figure 2-4B). The same residue exhibits severe exchange broadening that is significantly reduced upon Mg^{2+} binding (Figure 2-3B). In contrast, for G7 below the internal loop, which does not exhibit severe exchange broadening either in the absence or presence of Mg^{2+} (Figure 2-3B), good agreement was observed both in the absence (Figure 2-4A) and presence of Mg^{2+} (Figure 2-4B). Thus, Mg^{2+} binding appears to stabilize a Watson-Crick geometry for the G7-C30 base-pair primarily by affecting the alignment of C30. An excellent stem I order tensor fit (rmsd = 0.7 Hz) could be obtained in the presence of Mg^{2+} when all RDCs from the G7-C30 base-pair were included (Figure 2-4B, inset).

The only significant deviation in stem II was observed for terminal residue C9, both in the absence (Figure 2-4C) and presence of Mg^{2+} (Figure 2-4D). This deviation likely reflects an alternative conformation that accommodates the inter-helical kinking observed in SL1m (see below). Otherwise, an excellent fit was obtained for all remaining base-pairs in stem II (U10-A25 to G15-C20) both in the absence (Figure 2-4C, inset) and presence of Mg^{2+} (Figure 2-4D, inset). Thus, despite difficulty in observing imino signals for U10 and U11 in the absence of Mg^{2+}, the RDCs suggest that these residues adopt an A-form helical conformation both in the absence and presence of Mg^{2+}. Such an uninterrupted helix is consistent with conformer A but not conformer B (Figure 2-2C). While this suggests that the A conformer is the major SL1 species in solution both in the absence and presence of Mg^{2+}, this does not preclude dynamical averaging with the B conformer or other intermediates, especially if these species are populated to a low level (<10%) such that their contributions to the population weighted RDCs is negligible (20).

The observed chemical exchange broadening and RDCs suggest that Mg^{2+} arrests the dynamical exchange by stabilizing the A conformer.

Table 2-1: RDCs measured in SL1 in the absence and presence of 5 mM Mg^{2+}

Residue	Free (Hz)	+ Mg^{2+} (Hz)	Residue	Free (Hz)	+ Mg^{2+} (Hz)
2 (C8-H8)	24.6	N/A	20 (C5-H5)	32.8	27.8
3 (C1'-H1')	-18.9	N/A	20 (C6-H6)	17.7	16.4
4 (C1'-H1')	-17.9	N/A	20 (C1'-H1')	N/A	-12.3
5 (N3-H3)	-19.8	-13.4	21 (N1-H1)	-9.0	-8.1
5 (C5-H5)	30.5	24.5	21 (C8-H8)	13.8	14.2
6 (C5-H5)	N/A	28.54	22 (C8-H8)	N/A	10.2
7 (C8-H8)	26.5	19.4	22 (N1-H1)	-13.1	-8.6
8 (C8-H8)	24.3	13.8	24 (C2-H2)	28.9	19.6
9 (C5-H5)	28.9	N/A	24 (C1'-H1')	-26.2	N/A
9 (C6-H6)	14.8	N/A	24 (C8-H8)	32.0	24.7
9 (C1'-H1')	-34.7	-31.8	25 (C2-H2)	36.0	25.7
10 (C5-H5)	27.2	21.3	25 (C8-H8)	33.1	27.9
10 (C6-H6)	N/A	27.3	26 (C8-H8)	30.9	26.9
10 (N3-H3)	N/A	-12.4	26 (C1'-H1')	-10.7	N/A
11 (C5-H5)	30.6	28.0	27 (C2-H2)	33.2	28.7
11 (N3-H3)	N/A	-13.2	27 (C8-H8)	31.0	N/A
12 (N1-H1)	-14.9	-9.6	27 (C1'-H1')	0.8	-2.4
12 (C8-H8)	N/A	26.1	28 (C8-H8)	18.1	18.0
13 (C1'-H1')	N/A	-9.1	28 (C1'-H1')	-2.0	3.7
14 (C1'-H1')	N/A	-14.0	29 (C1'-H1')	-2.1	3.4
14 (C5-H5)	33.1	N/A	30 (C5-H5)	5.2	18.5
14 (N3-H3)	-6.5	-5.0	31 (C8-H8)	28.5	N/A
15 (C8-H8)	N/A	13.0	31 (N1-H1)	-18.6	-12.9
15 (N1-H1)	-8.5	N/A	33 (N1-H1)	-18.2	-12.5
16 (C8-H8)	17.9	10.2	32 (C2-H2)	30.7	21.8
16 (N1-H1)	9.7	9.9	32 (C8-H8)	39.6	26.6
17 (C2-H2)	28.2	28.2	33 (C8-H8)	36.1	28.6
17 (C8-H8)	17.8	20.4	34 (N3-H3)	-15.5	N/A
18 (C1'-H1')	-1.7	11.0	34 (C6-H6)	N/A	24.3
18 (C8-H8)	21.7	15.7	35 (C5-H5)	N/A	29.5
19 (C1'-H1')	19.5	15.2	35 (C6-H6)	N/A	24.4
19 (C2-H2)	24.3	17.9	35 (C1'-H1')	-10.1	N/A
19 (C8-H8)	29.5	16.7			

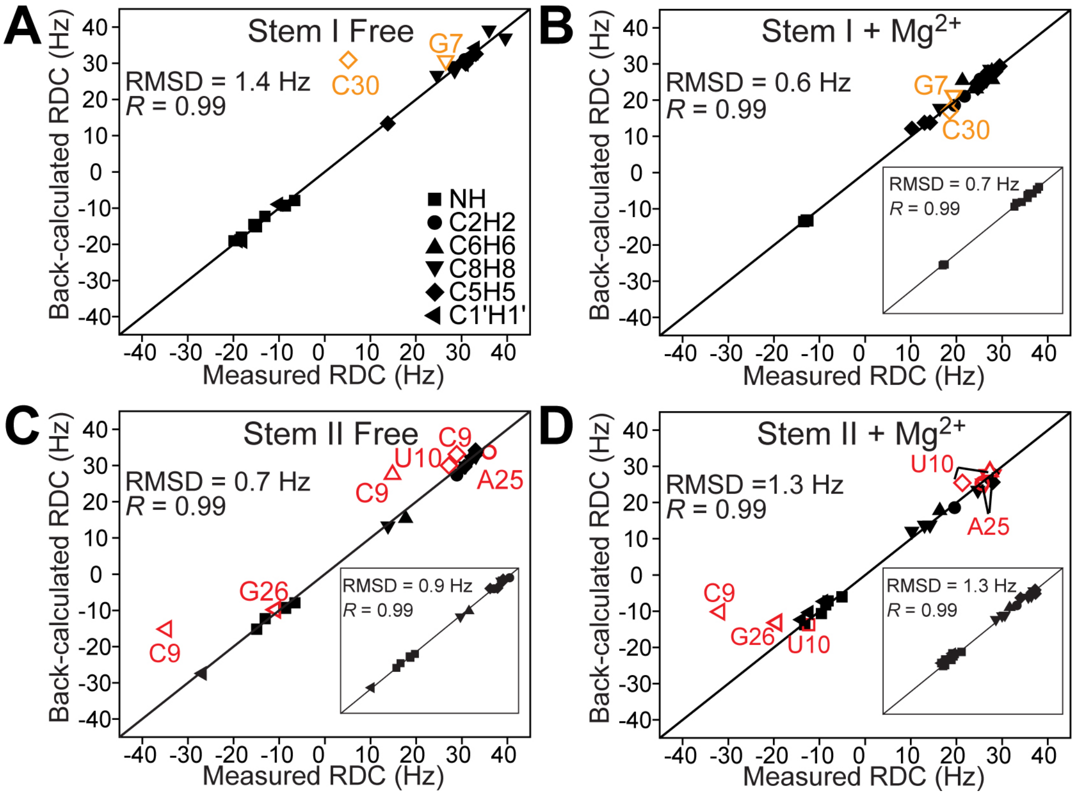

Figure 2-4. Order tensor analysis of RDCs measured in stem I in the (A) absence and (B) presence of Mg^{2+} and in stem II in the (C) absence and (D) presence of Mg^{2+}. Shown are correlation plots between measured and back-calculated RDCs when independently fitting stem order tensors to an idealized A-form geometry. Data that was excluded from the order tenser fit is shown in colored open symbols. The root-mean-square-deviation (RMSD) and correlation coefficient (*R*) is shown on each plot. The corresponding fits obtained when including all RDCs in the order tensor fit (excluding C9-G26 in stem II) are shown in the insets.

2.2.5 Mg^{2+} binding does not significantly affect the average internal loop conformation

The RDCs measured in the internal loop in the absence and presence of Mg^{2+} were very similar (Table 2-1). After accounting for the differences in degrees of order in the absence and presence of Mg^{2+} (Table 2-2), the rmsd between the free and Mg^{2+} RDCs in the internal loop is only 2.9 Hz, noting that comparable differences are expected due to differences in the orientation of overall alignment (S_{zz} direction differs by ~5°, data not shown). This indicates that the Mg^{2+} binding does not significantly alter the SL1m internal loop conformation.

Table 2-2: Order tensor analysis for RDCs measured in SL1m in the absence and presence of 5 mM Mg^{2+}. Shown are the number of measured RDCs (*N*), the condition number (*CN*) (12) describing the orientational spread of the RDC-targeted bond vectors, the root-mean-square deviation (rmsd) and correlation coefficient (*R*) between measured and back-predicted RDCs, the order tensor asymmetry ($\eta = |S_{yy}-S_{xx}|/S_{zz}$), generalized degree of order (ϑ), and internal generalized degree of order (ϑ_{int}).

SL1m	Stem	*N*	*CN*	RMSD (Hz)	*R*	η	$\vartheta \times 10^{-3}$	ϑ_{int}
Free	I	14	3.7	1.4	0.99	0.19±0.04	1.68±0.06	0.88±0.04
	II	17	2.9	0.9	0.99	0.14±0.04	1.47±0.03	
$+Mg^{2+}$	I	14	6.1	0.7	0.99	0.15±0.06	1.24±0.06	0.99±0.05
	II	23	2.8	1.3	0.99	0.15±0.04	1.23±0.03	

To obtain further insight into the internal loop conformation and possible variations upon Mg^{2+} binding, we used an RDC-based structure-validation approach in which we evaluated the agreement between the measured RDCs and internal loop conformations reported in three previous NMR structures (1N8X (2), 2GM0 (21), and 2D17 (22)) determined in the absence of divalent ions. In this analysis, non-terminal base-pairs in the A-form stem II were superimposed against the corresponding base-pairs in each NMR structure. The stem II order tensor determined in the absence and/or presence of Mg^{2+} (Table 2-2) was then used to predict free and Mg^{2+}-bound internal loop RDCs, respectively, using the NMR conformations as input coordinates and these were then compared to measured values.

In all three NMR structures, residues G8 and A27 form a G-A mismatch. Nevertheless, variations in the G-A geometry are observed across the NMR ensemble as reflected in the scatter of the predicted RDCs shown in Figure 2-5A. Many of the models in the three NMR structures reproduce both the free and Mg^{2+} RDCs. Interestingly, the Mg^{2+} RDC of G8 (C8H8) is observed to be in better agreement with models that differ from those that give good agreement with the corresponding free RDC with the Mg^{2+}

value favoring model structures in which the two bases are in a more co-planar alignment (e.g. models 13 and 14 in 2GM0 and models 5 in 2D17). The sugar C1'H1' A27 RDC was found to be in significantly better agreement with models from 2GM0, which generally yields the best agreement among the three structures examined.

The conformation of internal loop residues G28 and G29 which are implicated in NC binding (23-25) also varies significantly across the three structures as well as across various models within a given structure, particularly for 1N8X and 2D17 (Figure 2-5). For the G28 base, the best agreement was generally observed for conformations in which it is nearly co-planar with A27 (e.g. models 7 and 8 in 2D17, models 6 and 12 in 1N8X, and models 13 and 14 in 2GM0). Although no base RDC was available for G29, the best agreement with the corresponding C1'H1' RDC was observed for looped out conformations in many 2GM0 models as well as select 1N8X models (models 10, 14, and 18). Overall, the RDCs are consistent with conformations in which G29 is more extrahelical and accessible than G28 with no significant differences observed in the absence and presence of Mg^{2+}.

Finally, it is noteworthy that the C1'H1' RDCs measured in A27, G28 and G29 were all nearly zero both in the absence and presence of Mg^{2+} (Figure 2-5). Similar values were reported in the Mg^{2+} free duplex structure (21). While it was possible to find static conformations that can reproduce these RDCs, one cannot rule out the presence of extensive dynamical averaging which goes to attenuate the measured RDCs.

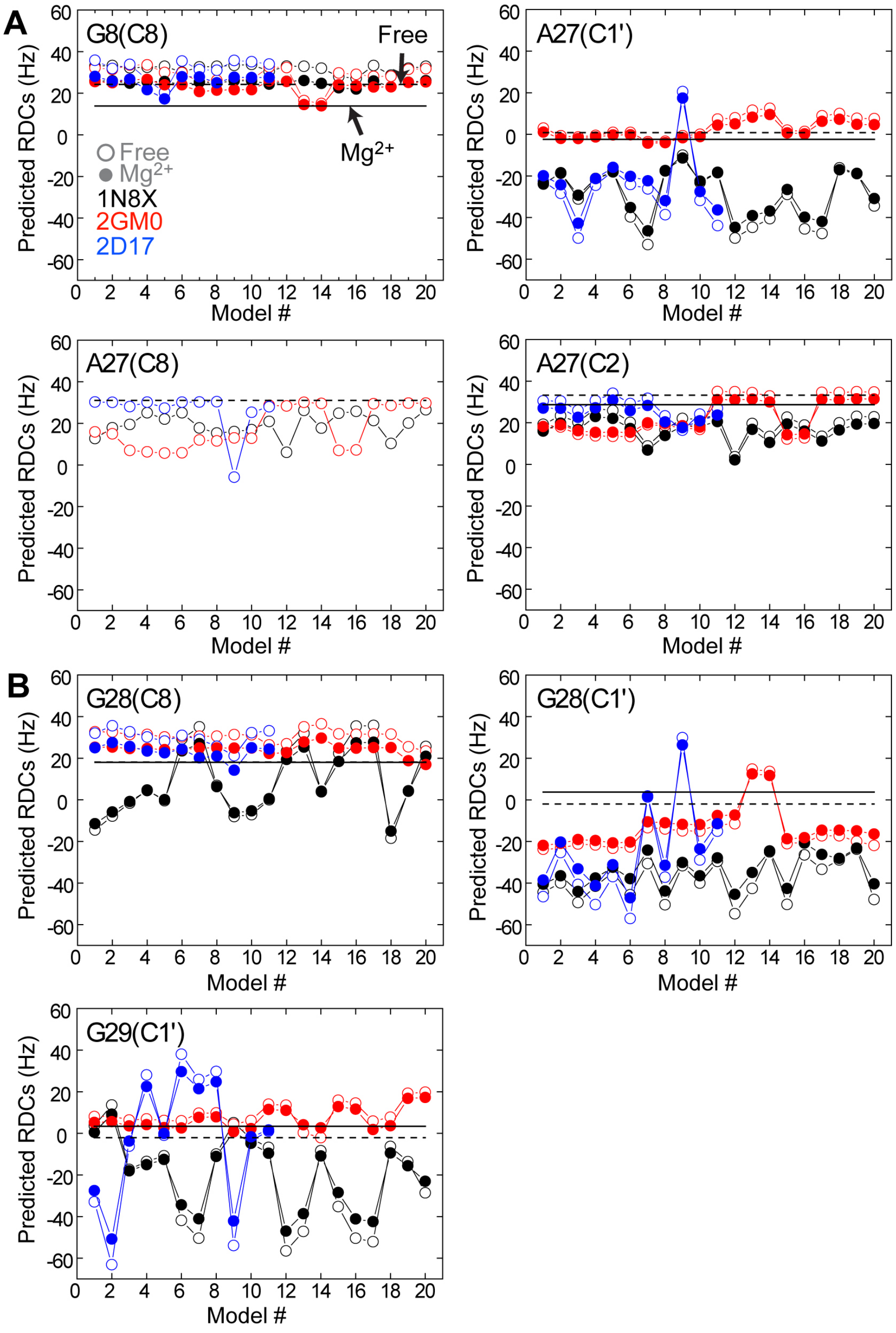

Figure 2-5. The conformation of the internal loop in the presence and absence of Mg^{2+} probed using RDCs. The RDCs measured in the absence and presence of Mg^{2+} are shown as dashed and straight horizontal lines, respectively. The free and Mg^{2+} RDCs predicted for a family of internal loop conformations in three different NMR structures of SL1 (1N8X (in black) (2); 2GM0 (in red) (21); 2D17 (in blue) (22)) are shown in open and filled symbols, respectively. For the duplex dimer structure 2GM0, every model yielded two symmetrically related internal loop conformations. The model numbering used corresponds to the following order; model 1 internal loop 1, model 1 internal loop 2, model 2 internal loop 1, model 2 internal loop 2, etc). For 2D17, the monomeric structure ensemble was used.

2.2.6 Mg^{2+} binding arrests inter-helical motions without significantly altering the average inter-helical alignment

The order tensors determined for stems I and II using the idealized A-form helix were used to determine the relative orientation and dynamics of the stems in the absence and presence of Mg^{2+}. Three order tensor elements define an order tensor frame (S_{xx}, S_{yy}, S_{zz}) that describes the average orientation of each stem relative to the applied magnetic field. The average relative orientation of stems can be determined by superimposing their respective order tensor frames (9,11-13). The inter-helical SL1m conformation determined in this manner in the absence and presence of Mg^{2+} is shown in Figure 2-6A and 2-6B, respectively. In both cases, the inter-helical angle only slightly deviates from coaxial alignment (10° and 14°, respectively). The differences between the bend/twist angles obtained in the absence (10° ± 4° / 8° ± 15°) (Figure 2-6A) and presence (14° ± 4° / 39° ± 15°) of Mg^{2+} is small and comparable to the experimental uncertainty which takes into account both the RDC uncertainty and A-form structural noise (Figure 2-6B) (10). Thus, as observed for the internal loop, Mg^{2+} binding does not significantly alter the average relative orientation of the two stems. The observed inter-helical bend/twist angles are in good agreement with the range of values reported in previous NMR structures of SL1 containing the same G-AGG internal loop (17°-27°/28°-65°, 14°-24°/24°-72°, 5°-42°/48°-78° for 1N8X, 2GM0, and 2D17, respectively). The small over twisting observed in the absence of Mg^{2+} is consistent with partial melting of base-pairs in stem II while the 33° over twisting observed in Mg^{2+} is consistent with “insertion” of a G-A base-pair between the two otherwise fully hydrogen bonded A-form stems.

We previously reported evidence for nanosecond stem motions in SL1m in the absence of Mg^{2+} using motionally decoupled ^{15}N spin relaxation data (4). The order tensor analysis also yields a generalized degree of order (GDO, ϑ) which can be used to assess motions between the two stems (12). Unlike spin relaxation data, which are used to probe internal motions at nanosecond and faster timescales, the GDO is sensitive to internal motions spanning a wider range of timescales (< millisecond). The GDO describes the degree of alignment experienced by each stem (12). While rigidly held stems should experience a common degree of alignment and therefore report identical GDOs, inter-stem motions can result in different degrees of stem order and hence stem GDOs (12). The stem GDO ratio, referred to as the internal generalized degree of order (GDO_{int}, ϑ_{int}), provides a measure of inter-stem motional amplitudes, ranging between 1 for inter-stem rigidity and 0 for maximum inter-stem motions (12).

In the absence of Mg^{2+}, the GDO computed for stem II (1.47×10^{-3} ± 6%) was smaller than that computed for stem I (1.68×10^{-3} ± 6%), yielding a GDO_{int} value of ~0.88 ± 4% (Figure 2-6A, Table 2-2). This indicates that the two stems dynamically reorient relative to one another in the absence of Mg^{2+}. Remarkably, the amplitude of inter-stem motions obtained using RDCs is in excellent agreement with that obtained independently by ^{15}N spin relaxation data ($S_s = 0.79^{1/2} = 0.89$) (4), indicating that the inter-stem motions occur at nanosecond timescales and that little to no inter-stem motions occur at micro-to-millisecond timescales. The larger degree of order observed for stem I is consistent with having the flexible hinge closer to stem II, which is expected given the partial melting of its three lower base-pairs. Assuming isotropic motions in a cone (13,26), the observed GDO_{int} value translates into inter-stem motional amplitudes of

~23°. This yields an SL1m conformation with the inter-helical angle that can vary dynamically between ~0° and ~33°. A similar range of bend angles (5°- 42°) is observed in the ensemble family of three previously reported NMR structures of SL1 containing the AGG internal loop (2,21,22).

In stark contrast, the stem GDOs obtained in the presence of Mg^{2+} are very similar ($1.24 \times 10^{-3} \pm 4\%$ and $1.23 \times 10^{-3} \pm 4\%$ for stems I and II, respectively), yielding a GDO_{int} value that is close to unity (0.99 ± 5%) (Figure 2-6B, Table 2-2). This suggests that while Mg^{2+} binding does not significantly affect the average orientation of the two stems, it arrests their relative motions, stabilizing the global SL1m conformation.

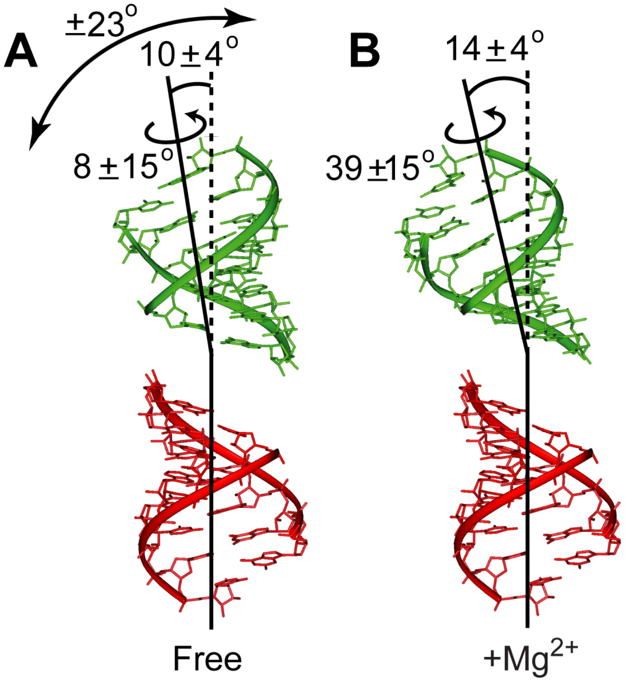

Figure 2-6. Relative orientation and dynamics of the two SL1m stems in (A) the absence and (B) presence of Mg^{2+} derived from order tensor analysis of the stem RDCs. Shown are the values of the inter-helical bend and twist angles together with the amplitude of internal motions obtained from the ratio of stem GDOs assuming an isotropic cone motional model.

2.2.7 Collective and librational motions in the absence and presence of Mg^{2+} from dynamically decoupled ^{15}N relaxation data

The above RDC analysis can be compromised by correlations between stem motions and overall alignment which are expected given the similar size and shape of the two SL1m stems. For example, an alternative interpretation for the apparent arrest of stem motions by Mg^{2+} could be that by stabilizing the three lower base-pairs in stem II, Mg^{2+} equalizes the contribution of each stem to the total alignment in which case similar stem GDOs are expected regardless of inter-stem motions (27).

To confirm that Mg^{2+} binding arrests stem motions, we used a recently described domain-elongation strategy for decoupling stem motions from overall reorientation (4). The elongation slows down overall molecular tumbling, allowing resolution of motions that may otherwise evade detection because they occur at timescales approaching overall rotational diffusion. Elongation by twenty-two base-pairs is carried out using two constructs which are designed to reduce spectral overlap due to elongation residues (4). Using imino ^{15}N relaxation data, we previously resolved stem motions in elongated SL1m (E-SL1m) in the absence of Mg^{2+} that evade detection in non-elongated SL1m (4). Thus, while the transverse (R_2) to longitudinal (R_1) ^{15}N relaxation rate ratios (R_2/R_1) measured in stems I and II were similar in SL1m (4), those of stem II were significantly attenuated relative to stem I in E-SL1m (Figure 2-7C). Extended Model-Free analysis (28,29) of the data yielded time constants for overall rotational diffusion (τ_M) and both amplitudes and time constants for fast local (S_f^2/τ_f) and slow stem (S_s^2/τ_s) motions (Figure 2-7D). Here, S^2 is the Lipari-Szabo spin relaxation order parameter (29) which varies between 0 and 1 for maximum and minimum motions, respectively.

In the previous study of E-SL1m, we used spectral comparisons such as those shown in Figure 2-7B to show that elongation does not disrupt the structural and/or dynamical integrity of SL1m. The similar exchange-broadening pattern (Figure 2-8A) and Mg^{2+} induced chemical shift changes (Figure 2-7B and 2-3B) observed in SL1m and E-SL1m further suggest that elongation does not affect the dynamical and Mg^{2+} binding properties of SL1m. To this end, we measured ^{15}N relaxation data (Table 2-3 and 2-4) in E-AU-SL1m and E-GC-SL1m (Figure 2-7A) in the presence of 10 mM Mg^{2+}. Assuming two-state binding with a K_d ~ 0.9 mM (Figure 2-3C), up to 84% of E-SL1m is expected be in the Mg^{2+} "bound" state.

Table 2-3: ^{15}N relaxation data measured in E-SL1m in the presence of 10 mM Mg^{2+}

Residue	R_2 (Hz)	R_1 (Hz)	NOE
U5 (N3)	31.51±0.37	0.645±0.018	0.672±0.030
G31 (N1)	30.61±0.37	0.671±0.011	0.670±0.032
G33 (N1)	30.06±0.42	0.718±0.013	0.636±0.056
U10 (N3)	29.40±0.53	0.688±0.023	0.701±0.036
U11 (N3)	28.54±0.19	0.739±0.016	0.652±0.028
U14 (N3)	32.66±0.41	0.734±0.018	0.591±0.023
G15 (N1)	27.02±0.47	0.813±0.012	0.524±0.039
G16 (N1)	27.93±0.81	0.827±0.024	0.759±0.084
G21 (N1)	31.01±0.31	0.823±0.013	0.565±0.038
G22 (N1)	26.89±0.61	0.733±0.008	0.631±0.042

As shown in 2-7E, Mg^{2+} binding led to a significant increase in the stem II R_2/R_1 values without significantly affecting the stem I values (Figure 2-7E). The resulting R_2/R_1

values measured in stems I and II are far for more similar in the presence of Mg^{2+} (Figure 2-7E). This is exactly as would be expected if Mg^{2+} binding caused a reduction in the amplitudes of inter-stem motions. The stem II R_2/R_1 values remain smaller than stem I and/or values predicted assuming a static RDC-derived SL1m+Mg^{2+} conformation (Figure 2-7E, open symbols). This is not surprising given that ~16% of E-SL1m is expected to be on average in the dynamical free state. Interestingly, very good agreement is observed between the measured stem II R_2/R_1 values and those back-predicted assuming a population weighted average of free (measured experimentally) and Mg^{2+} bound (computed hydrodynamically) R_2/R_1 values (Figure 2-7E, in grey).

The ^{15}N relaxation data was analyzed using the extended Model-Free analysis (28,29). This analysis was repeated for a range of input E-SL1m conformations that reflect the uncertainty in the inter-stem alignment obtained by RDCs. With the exception of loop residue G16, for which the dynamical parameters obtained varied considerably depending on the input conformation, consistent dynamical parameters were obtained for all stem base-pairs (data not shown). The dynamical parameters obtained using the best-fit E-SL1m+Mg^{2+} RDC-derived conformation is shown in Figure 2-7D and Table 2-4. The time constant for overall rotational diffusion ($\tau_M = 22.3 \pm 0.1$ and 23.5 ± 0.1 ns for E-AU-SL1m and E-GC-SL1m, respectively) is in very good agreement with values predicted hydrodynamically (22.2 ns) (30). Relative to free E-SL1m (Figure 2-7D), Mg^{2+} binding leads to a uniform increase in the S_f^2 values and thus decrease in the local N-H librations, suggesting local stabilization of the two stems (Figure 2-7F). A larger uniform increase in the S_s^2 values of stem II is also observed indicating that Mg^{2+} binding causes a reduction in the amplitudes of stem motions (Figure 2-7F). Though small changes in the time constants for librations (τ_f) and stem motions (τ_s) are also observed upon Mg^{2+} binding, comparable variations were observed when using distinct input SL1m conformations that reflect the RDC measurement uncertainty (data not shown).

Table 2-4: Dynamical parameters for E-SL1m in the presence of 10 mM Mg^{2+} obtained from Model-Free analysis of the ^{15}N relaxation data and the RDC derived mean inter-stem orientation.

Residue	Model	S^2_f	S^2_s	$_f$(ps)	$_s$(ns)	$_i$	$_m$(ns)
G15	5	0.935±0.026	0.887±0.017		0.92±0.20	0.00	22.3±0.050
G16	5	0.820±0.077	0.734±0.124		4.16±7.16	0.00	
G21	5	0.906±0.018	0.931±0.015		0.49±0.21	0.00	
G22	5	0.865±0.029	0.938±0.019		0.82±0.35	0.00	
G31	2	0.943±0.019		68.023±36.152		0.00	
G33	1	0.910±0.020				2.50	
U5	2	0.981±0.015		244.504±852.469		0.42	23.5±0.056
U10	1	0.919±0.029				3.03	
U11	5	0.947±0.016	0.931±0.014		1.21±0.33	0.00	
U14	5	0.858±0.020	0.864±0.016		1.29±0.16	0.00	

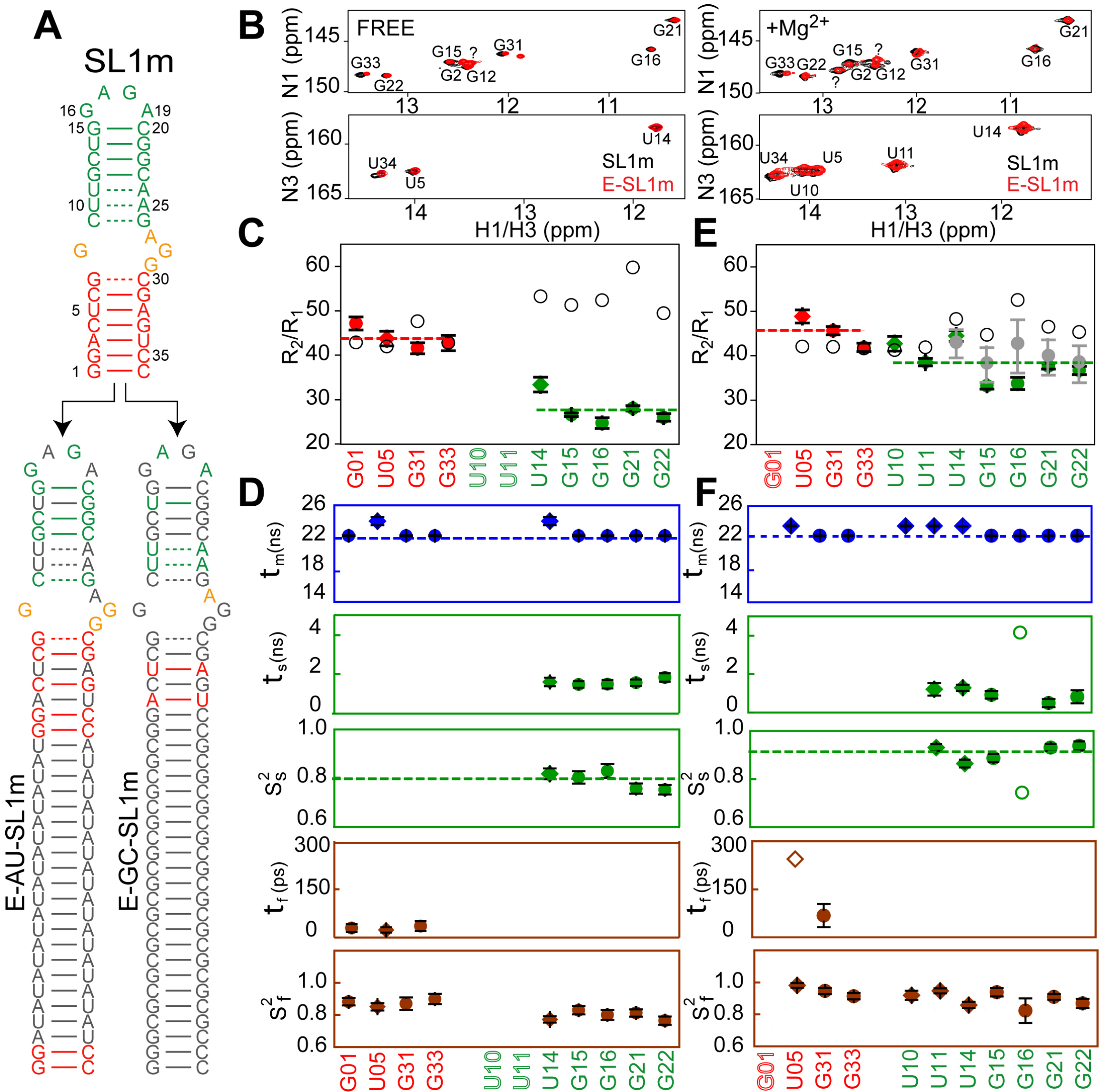

Figure 2-7. Extended Model-Free analysis of ^{15}N relaxation data measured in E-SL1m in absence (4) and presence of Mg^{2+}. (A) Secondary structure of E-GC-SL1m and E-AU-SL1m. ^{13}C/^{15}N labeled and unlabeled residues are shown in color and gray, respectively. (B) 2D ^{1}H-^{15}N HSQC spectra of the imino region of E-GC-SL1m + E-AU-SL1m (black) overlaid on non-elongated SL1m (red) in the absence and presence of Mg^{2+}. Watson-Crick guanine signals that could not be directly assigned using NOESY data are labeled with '?' and may correspond to either G26 or G7. (C-D) E-SL1m dynamics in the absence of Mg^{2+} (4). (C) R_2/R_1 values measured in guanine (filled circles) and uridine (filled diamonds) residues in stems I (red) and II (green) in the absence of Mg^{2+}. Values predicted using hydrodynamic calculations are shown as black open circles. Error bars represent the measurement uncertainty. (D) Dynamical parameters for E-SL1m in the absence of Mg^{2+} obtained previously (4) from an extended Model-Free analysis (28,29). Shown are the time constants/amplitudes for fast (τ_f / $S^2{}_f$) and slow (τ_s /$S^2{}_s$) internal motions. (E-F) Dynamics in E-SL1m in the presence of 10 mM Mg^{2+}. (E) Measured R_2/R_1 values (filled color symbols) and those predicted hydrodynamically assuming a

static conformation (black open circles) and a population weighted average of free and Mg^{2+} bound states (see text) (filled gray circles). The error bar in the latter reflects different relative populations of free and bound states due to the spread of the observed K_ds. (F) Dynamical parameters obtained in the presence of 10 mM Mg^{2+} using extended Model-Free analysis of the ^{15}N relaxation data. Results for loop residue G16 (open symbol) varied considerably when using different input E-SL1m+Mg^{2+} conformations that reflect the RDC uncertainty and is therefore considered to be unreliable.

2.2.8 Local dynamics of internal loop in G28 and G29 in the presence of Mg^{2+}

In our study of free E-SL1m, the C-H resonance intensities exposed internal motions that evade detection in non-elongated SL1m because they occur at timescales approaching overall molecular tumbling (4). Ignoring chemical exchange, the intensities of non-exchangeable C-H resonances reflect the reorientation of dipolar and chemical shift anisotropy (CSA) tensors relative to the applied magnetic field. Placement of the base dipolar and ^{13}C CSA tensors nearly perpendicular to the long axis of diffusion renders the base stem I intensities near maximum values expected from anisotropic tumbling of a rigid molecule (4). Thus, resonances that exhibit higher intensities correspond to regions with added-on internal mobility occurring at timescales faster than overall molecular tumbling.

To gain further insight into how Mg^{2+} binding affects the dynamics of SL1m, we measured resonance intensities in SL1m and E-SL1m in the presence of Mg^{2+} at 25°C and compared results with intensities observed in the free state (Figure 2-8). For free SL1m, small variations in intensities are observed which are reduced upon Mg^{2+} binding (Figure 2-8B). In contrast, due to the resolving power afforded by elongation, the intensities in free E-SL1m exhibit much wider variations from site to site that are indicative of variable degrees of internal motions (4). The changes in intensities following Mg^{2+} binding are also far more significant in E-SL1m compared to SL1m

(Figure 2-8D). In addition to a decrease in exchange broadening in and around the internal loop, the stem II and GAGA tetraloop intensities experience a significant reduction, consistent with a long-range arrest of inter-stem motions. A significant intensity reduction is also observed for A27, possibly reflecting stabilization of the G8-A27 mismatch, though new exchange broadening contributions cannot be ruled out. Remarkably, and in stark contrast, a significant increase in the intensities is observed for the base moieties of G29 and to a smaller extent G28 (Figure 2-8D). This suggests that Mg^{2+} binding activates nanosecond local motions in the base moieties of G29 and to lesser extent G28. However, it is also possible that these motions were present in free E-SL1m but they were masked by exchange-broadening contributions that are reduced upon Mg^{2+} binding. Though spectral overlap in non-elongated SL1m did not permit measurement of the corresponding resonance intensity or RDC for G29, elevated intensities are not observed for G28 in SL1m (Figure 2-8B) most likely because the motions occur at timescales approaching overall molecular tumbling. The high mobility observed at G29 and G28 in the presence of Mg^{2+} is significant given that these are precisely the internal loop residues that are implicated in NC binding (23-25).

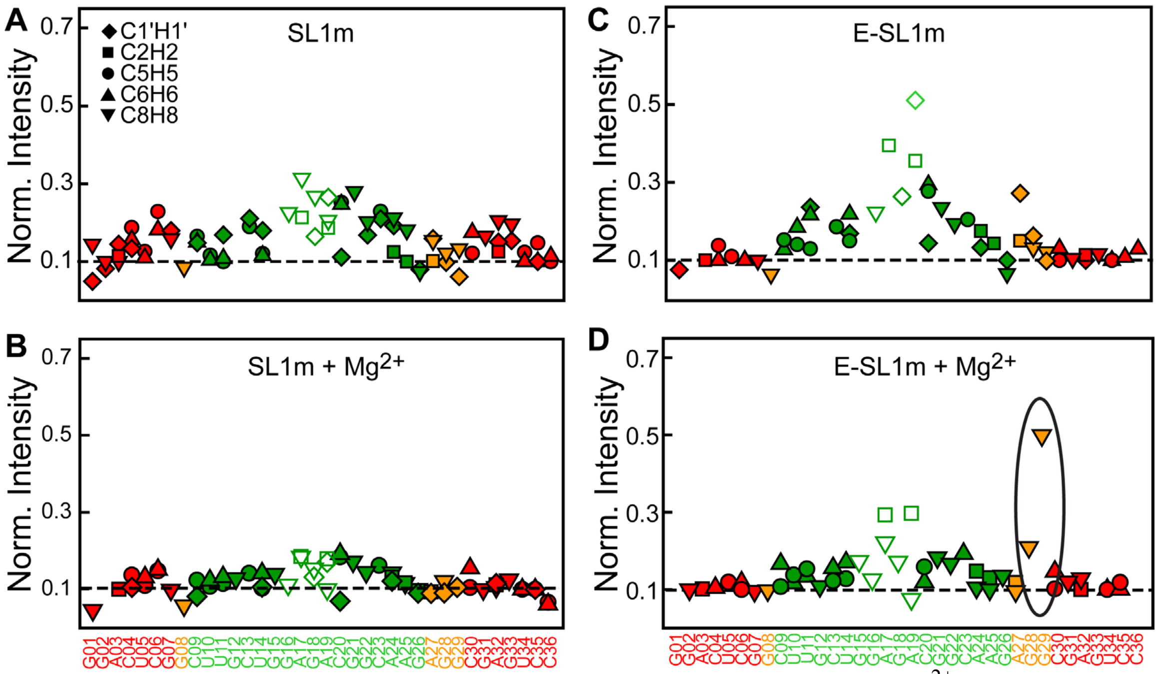

Figure 2-8. Dynamics in E-SL1m in the absence and presence of Mg^{2+} determined by motional narrowing of resonances. Shown are the normalized resonance intensities as a function of residue in the presence and absence of Mg^{2+} for (A, B) SL1m and (C, D) E-SL1m at 25°C. Resonance intensities are obtained from ^{1}H-^{13}C HSQC experiments. The intensity for a given type of C-H vector is normalized to a minimum nominal value of 0.1 indicated by a horizontal line. Peaks that exhibit exchange broadening as inferred from the temperature-induced perturbations were not used in the above normalization and generally have intensities < 0.1.

2.2.9 Neomycin B binds the SL1 internal loop and neighboring residues.

Although neomycin B has already been shown to inhibit the viral production of HIV-1 up to 85% in a dose dependent fashion in U1 cells (31), its mode of action and primary HIV RNA target(s) remains to be fully established. It is known that neomycin B can bind multiple regulatory HIV-1 RNA elements. Foot printing experiments on a 171mer RNA derived from the packaging region of the HIV-1 RNA genome indicated that neomycin B and paromomycin bind the SL1 stem-loop in a region involving the helical domains and the internal loop (32,33). In addition, the crystal structure of the neomycin and SL1 kissing hairpin complex revealed that two neomycin molecules interacted with the unpaired A on the each side of the apical loop in the kissing dimer. To further characterize the SL1-neomycin B interaction, we preformed chemical shift titrations of the mutant SL1m with neomycin B. As highlighted in spectra of Figure 2-3D, neomycin B leads to specific chemical shift perturbations which are the same as in the case of Mg^{2+}(Figure 2-3B). In the presence of neomycin B, the first group of resonances to disappear had the largest chemical shift perturbations in the presence of Mg^{2+}. This was particularly pronounced for the internal loop residues and neighboring helical residues G8, G9, G26, and G28. The rest of resonances, such as G7, U10, and G31, shifted towards the same direction in both. The resonances exhibit a combination of fast and intermediate NMR exchange kinetics, indicating relatively tight binding. These results argue that neomycin B binds the internal loop and neighboring residues in SL1, potentially inducing a change in its conformation and that the SL1-neomycin B complex can be targeted for NMR investigation.

2.3 Discussion

SL1 is a highly conserved stem-loop in the HIV-1 leader RNA that is believed to be involved in functionally important structural transitions that are modulated by Mg^{2+} binding and that are catalyzed by NC. In our study, we examined the dynamical and Mg^{2+} binding properties of the SL1 structure with the goal of obtaining new insight into the molecular basis of its structural transitions.

Our study suggests that the highly conserved SL1 internal loop sequence-specifically destabilizes the upper stem by allowing formation of two (or possibly more) competing secondary structures (Figure 2-2C). This conformational equilibrium is intricately dependent on the sequence of the GGA internal loop and neighboring residues in stem II. For example, the reduced exchange broadening observed in the GGA internal loop mutant can be explained by its inability to slip into the B conformer (1). A similar argument could be used to explain the higher stabilities of other SL1 internal loop mutants (1). Interestingly, a uridine-substituted SL1 internal loop does not disrupt dimer formation but results in a mutant virus with diminished genome packaging (24). In addition to possibly interfering with NC binding, the uridine substitution is expected to impair formation of the B conformer and thus stabilize the internal loop and upper stem.

SL1 kissing dimers containing the internal loop can spontaneously convert into duplex dimers at 55°C in the absence of Mg^{2+} whereas constructs lacking the internal loop cannot (34). Our results show that the internal loop introduces internal flexibility into the SL1m structure that can promote the kissing-duplex transition. The transition requires the exchange of strands between monomers in the kissing dimer which in turn requires that strands from the two monomers come into close proximity. The two stems above the internal loop are likely candidates for initiating strand exchange since they are the most closely positioned in the kissing dimer. Studies have shown that the kissing-duplex transition can occur without disrupting the loop-loop interaction (34-36). The two monomers can be brought into close proximity without disrupting the loop-loop interaction by rotating each monomer around a direction perpendicular to the C2 axis of symmetry. Such a rotation ensures that C2 symmetry is maintained in the dimer during the transition. This leads to formation of an intermediate in which base-pairs in the upper stem are proximate and poised to form both inter- and intra-molecular hydrogen bonding. Such an intermediate has previously been proposed (37) and recently visualized by molecular dynamics simulations in the context of short kissing SL1 dimers lacking the internal loop (35). The inclusion of the internal loop is expected to destabilize the upper stem, making its base-pairs a key nucleation site for initiating the melting, exchange and reannealing of strands. This would explain why constructs lacking the internal loop cannot undergo the kissing-duplex transition spontaneously (34). Due to inter-helical kinking, the lower stems are not expected to be in immediate register in the intermediate, and inter-stem flexibility may play a role in bringing the lower stems into proper register for completing strand exchange.

Our results show that Mg^{2+} binds to the SL1 internal loop region without significantly altering the average SL1 structure. Electrostatic calculations on the SL1m structure determined in the absence of Mg^{2+} (2) show that residues that experience the largest Mg^{2+} induced chemical shift changes belong to regions with strong negative electrostatic potential. The RDCs and relaxation data do however show that Mg^{2+} arrests the dynamical equilibrium, stabilizing the hydrogen bond alignments in the upper stem while simultaneously reducing inter-stem motions. In this regard, Mg^{2+} binding is expected to reduce the likelihood for spontaneous kissing-duplex transitions in internal loop-containing SL1 constructs. So far, only one study has explored the effects of Mg^{2+} binding on the spontaneous transition in internal loop-containing SL1. Using a ribozyme-based cleavage assay to monitor the kissing-duplex transition, this study showed that increasing the Mg^{2+} concentration (up to 100 mM) resulted in an increase in the probability for spontaneous kissing-duplex conversion (38). However, at the low RNA concentrations used (~0.01 mM compared to ~1 mM used in other studies) the encounter of monomers may be rate-limiting and Mg^{2+} may increase the probability for encounter by stabilizing the loop-loop interaction (37).

In vivo, the transition between kissing and duplex dimers as well as between the LDI and BMH conformers is believed to be catalyzed by NC protein which is known to bind to exposed guanine residues (39,40). Previous mapping studies have shown that G28 and G29 are accessible in the context of the 206 nt Ψ HIV-1 RNA (25). Footprinting data on the 401 nt fragment of the RNA leader show that G28 and G29 are both strongly accessible in the free RNA and strongly protected upon binding to Gag and NC (41). A recent study suggests that NC binding to the SL1 internal loop is largely responsible for driving the kissing-duplex transition, whereas binding at the apical loop inhibits dimer formation (42). The latter study also showed that NC binding to the SL1 internal loop is similar whether in the monomer or kissing dimer context, further validating the relevance of our dynamical results on monomeric SL1m (42). Remarkably, our results show that while Mg^{2+} causes stabilization of the internal loop and upper stem, it increases or retains significant local flexibility in the base moieties of G28 and G29, likely making them available for NC binding. The RDCs measured in the presence and absence of Mg^{2+} are also consistent with an extra-helical conformation for G29 (21). In this manner, Mg^{2+} binding may help ensure that the kissing-duplex conversion does not occur spontaneously and prematurely but rather only following NC binding. The conformational pathway may involve the ejection of stabilizing Mg^{2+} ions by the basic region of NC which has been shown to be sufficient for catalyzing the dynamical transition (Figure 2-9) (43).

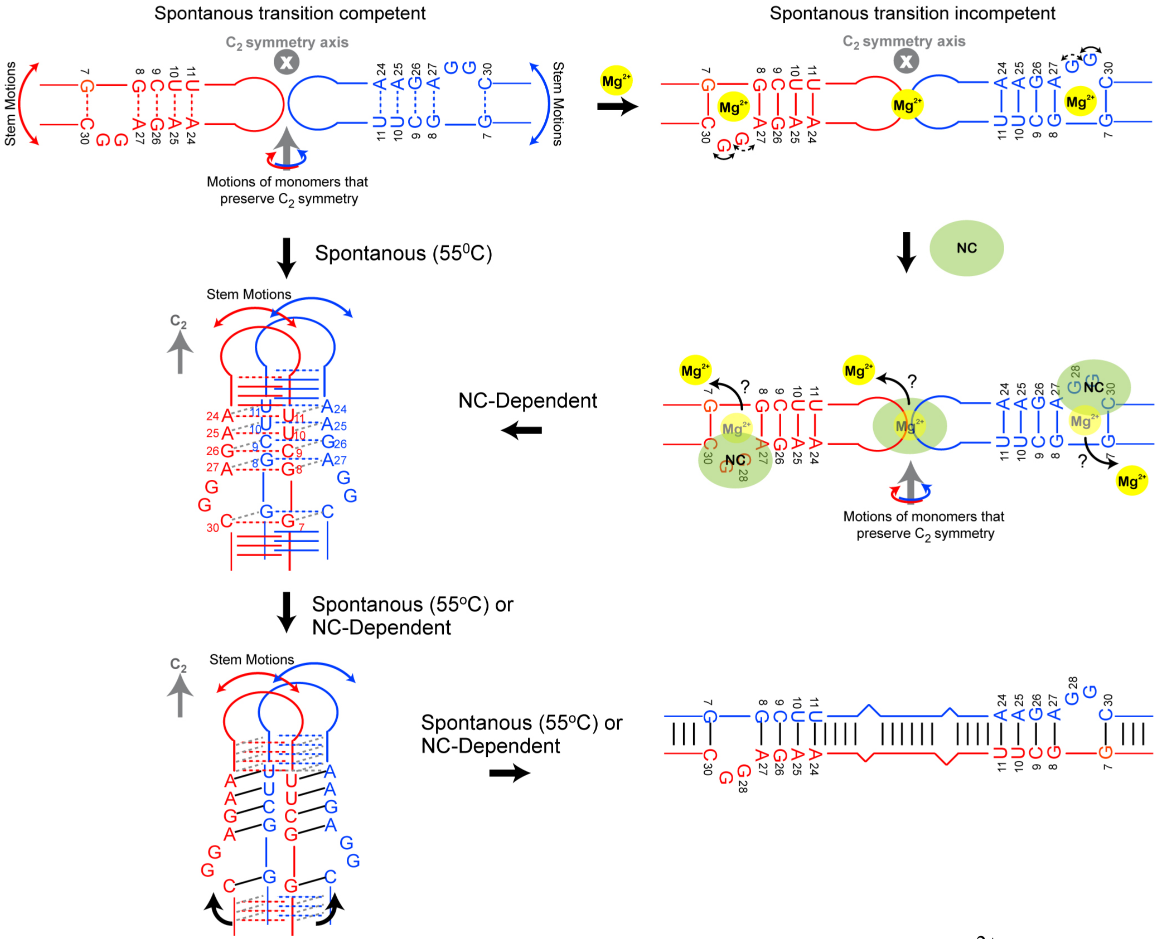

Figure 2-9. Proposed role for internal loop induced dynamics and Mg^{2+} binding in spontaneous and NC protein dependent SL1 structural transition between kissing and duplex dimers.

2.4 Materials and methods

2.4.1 Sample preparation

Uniformly $^{13}C/^{15}N$ labeled SL1m, SL1, and elongated SL1m (E-GC-SL1m and E-AU-SL1m, see Figure 2-7A) NMR samples (0.6-1.0 mM) were prepared using standard *in vitro* transcription reactions as previously described (4). The buffer used in NMR experiments contained 15 mM sodium phosphate, 25 mM sodium chloride, 0.1 mM EDTA, and pH ~ 6.4 in 8% $^{2}H_2O$. NMR sample for the measurement of RDCs contained

~15 mg/ml of Pf1 phage (Asla Biotech) (7) (8). The kissing SL1 dimer was prepared following the previously described procedure (34).

2.4.2 NMR spectroscopy

All NMR experiments were performed on an Avance Bruker 600 MHz NMR spectrometer equipped with a triple-resonance 5 mm cryogenic probe. Unless stated otherwise, all NMR experiments were conducted at 298 K. NMR spectra were processed using NMRPipe/NMRDraw (44), analyzed using NMRView (45), and overlaid using SPARKY 3 (46). All non-exchangeable resonances were assigned using standard methods (47) and found to be in excellent agreement with assignments reported previously for the same construct (2). With the exception of terminal residues G1, G7, and G26, all expected exchangeable imino signals belonging to Watson-Crick base-pairs (G2, U34, G33, U5, G31, U10, U11, G12, G22, U14, G21 and G15) could be assigned using the 3D ^{1}H-^{15}N HSQC-NOESY experiment in the absence or presence of Mg^{2+}. In the absence of Mg^{2+}, the U10 imino signal could not be observed, while that of U11 could only be observed at low temperatures (5°C) (Figure 2-3). Watson-Crick hydrogen-bond alignments were also detected using the J_{NN}-COSY experiment recorded at 5°C and 25°C (5) (6). *Trans*-hydrogen bond connectivity for U10 and U11 could only be established in the presence of Mg^{2+}. As in previous NMR studies (21), additional weak guanine signals and one uridine signal were observed that could not be unambiguously assigned. One of those peaks has a weak NOESY cross peak to G2, consistent with a G1 assignment. Through comparison with spectra of elongated SL1m, one uridine and two guanine signals could also be assigned to alternative forms of terminal residues G1, G2 and U34.

All RDCs were measured using Pf1 phage as an ordering medium (~15 mg/ml) (7) (8). One bond C-H RDCs ($^1D_{\mathrm{C1'H1'}}$, $^1D_{\mathrm{C2H2}}$, $^1D_{\mathrm{C5H5}}$, $^1D_{\mathrm{C6H6}}$, $^1D_{\mathrm{C8H8}}$) were measured in duplicate using 2D ^{13}C–^{1}H S^{3}CT-HSQC experiments (48,49) in which the C-H doublet is encoded along either the direct (^{1}H) or indirect (^{13}C) dimension. The two sets of RDCs yielded a standard deviation of 1.8 Hz for both free and Mg^{2+} bound SL1m, respectively. The average RDC value from the two sets was used whenever possible. One bond N-H RDCs ($^1D_{\mathrm{N1H1}}$ and $^1D_{\mathrm{N3H3}}$) were measured in duplicate using 2D HSQC experiments without ^{1}H or ^{15}N decoupling in the indirect (^{15}N) and direct (^{1}H) dimension, respectively. A total of 52 RDCs were measured in SL1m in the absence and presence of Mg^{2+} (Table 2-1). In all cases, spectra of SL1m in the presence and absence of Pf1 phage were virtually identical, both in the absence and presence of Mg^{2+} (data not shown).

Imino ^{15}N longitudinal (R_1) and transverse ($R_{2(\mathrm{CPMG})}$) relaxation rates and ^{1}H-^{15}N NOEs were measured in E-SL1m in the presence of 10 mM Mg^{2+} using well-established 2D experiments closely following the procedure described previously for free E-SL1m (4). The R_1 relaxation delays were 0.06, 0.12, 0.24, 0.48, 0.64, 0.80, and 1.2 s(50). The $R_{2(\mathrm{CPMG})}$ experiment employed a recently described $[0013]^N$ phase cycle in the Carr-Purcell-Meiboom-Gill (CPMG) pulse train for suppressing artifacts arising from off-resonance effects (51). The relaxation delays were 0.0062, 0.0124, 0.0248, 0.0372, 0.0496, 0.0620, and 0.0744 s. The final R_2 values were corrected to account for off-resonance effects as previously described (51). Uncertainties were also verified based on duplicate measurements of relaxation data. The ^{1}H–^{15}N NOEs were measured from pairs of spectra recorded with and without proton saturation during the recycle delay. The NOE uncertainty was estimated from the standard deviation of the baseline in the spectra. Resonance intensities were measured in E-AU-SL1m and E-GC-SL1m as previously described (4). Mg^{2+} chemical shift titrations were performed by recording 2D ^{1}H-^{13}C or ^{1}H-^{15}N HSQC spectra following incremental increases in the $MgCl_2$ concentration (0.1, 0.2, 0.4, 1.6, 3.2 and 5.0 mM).

2.4.3 Data analysis

The RDCs were analyzed using the order tensor formalism (9,13,52) tailored for A-form helices as implemented in the program Aform-RDC (10). Watson-Crick residues in stems I and II were modeled using idealized A-form helices constructed using the program Insight II (Molecular Simulations, Inc). The propeller twist angles had to be corrected from +15° to the standard A-form value of -15° (10). The 14/18 and 14/23 RDCs measured in non-terminal Watson-Crick base-pairs of stems I/II in the absence and presence of 5 mM Mg^{2+}, respectively, together with the idealized A-form helix geometry were used to obtain best-fit order tensors for each stem using singular value decomposition implemented in the program RAMAH (53) (an in-house modified version of ORDERTEN_SVD) (52). Statistics for the order tensor analysis are shown in Table 2-2. The relative orientation of the two stems was determined by superimposing their respective order tensor frames (52). Three of the four degenerate domain-domain orientations involving rotation of 180° about the three order tensor principal axes could be omitted due to structural constraints (54). The translational disposition of stems was determined approximately by modeling in the connector internal loop from the previous

NMR structure of the same GAGA SL1m mutant (2). The assembled SL1m structures were further validated using the program PALES (http://spin.niddk.nih.gov/bax/) (55). The S_{zz} orientations predicted using PALES based on the free and Mg^{2+} bound SL1m structures were in excellent agreement with that determined experimentally (differences are ~1° and ~3°, respectively). Inter-helical angles were calculated using in-house written software (27). The uncertainty in the order tensor parameters was computed using the program Aform-RDC, which takes into account both RDC uncertainty and parameterized structural noise in the idealized A-form geometry (10). An indirect order tensor fit was also carried out for internal loop RDCs using the previously reported NMR conformations (1N8X (2); 2GM0(21); 2D17(22)). In this analysis, central base-pairs in stem II (C13-G22, G12-C23, U11-A24, and U10-A25) were used to superimpose the idealized stem II helix against each NMR structure. Next, the best-fit order tensor determined for stem II (Table 2-1) was used to back-predict internal loop RDCs for each conformation. This analysis was carried out on each model in a given NMR ensemble.

The R_1 and R_2 values and their uncertainties measured from ^{15}N relaxation data were determined by non-linear least squares fitting of experimental data to the monoexponential function, $I(t)= I(0)e^{(-R\times t)}$ based on the Levenberg-Marquardt (LM) algorithm implemented in the software Origin (Origin Lab Corporation)(Figure 2-5). The R_1, R_2, and NOE values measured in E-SL1m+Mg were subjected to an extended Model Free analysis (28,29) using the software Modelfree (Version 4.16 for Linux) provided by Palmer and co-workers (56) closely following the procedure described previously (4). This analysis assumed (i) the decoupling approximation (28,29), (ii) an axially symmetric overall diffusion tensor, (iii) a diffusion tensor with D_{ratio} ($D_{//}/D_{\perp}$= 5.8) and orientation computed using HYDRONMR (30) and the RDC derived average SL1m+Mg^{2+} inter-stem conformation. Model selection was carried out following the procedure recommended by Palmer and co-workers (56) as recently described in detail (4).

Changes in chemical shift upon addition of Mg^{2+} were quantified using, $\Delta\delta = \sqrt{(\Delta\delta_H)^2 + (\alpha\Delta\delta_X)^2}$, where $\Delta\delta_H$ and $\Delta\delta_X$ are the chemical changes in the ^{1}H and ^{13}C/^{15}N dimensions, α is the ratio of the H and C/N gyromagnetic ratio. The apparent K_ds were by obtained fitting to the following equation (57):

$$\delta_{obs} = \frac{\Delta\delta_T\{([Mg^{2+}]_T + [RNA]_T + K_d) - \sqrt{([Mg^{2+}]_T + [RNA]_T + K_d)^2 + (4[Mg^{2+}]_T[RNA]_T)}\}}{2[RNA]_T}$$

where δ_{obs} is the observed chemical shift, $\Delta\delta_T$ is the chemical shift difference between the free and fully bound state, $[Mg^{2+}]_T$ is the amount of added $MgCl_2$, and $[RNA]_T$ is the RNA concentration obtained from its absorbance at 260 nm. The data was fitted using Origin (Origin Lab Corporation) in which the unknown values $\Delta\delta_T$ and K_d are optimized during the fit. Representative fits are shown in Figure 2-3C.

2.4.4 Computational modeling

Insight II (Molecular Simulations, Inc.) was used to *in silico* elongate stem I in the RDC derived SL1m+Mg^{2+} structure using a stretch of twenty-two base pairs. The resulting E-SLm+Mg^{2+} structure was subjected to hydrodynamic calculations using the program HYDRONMR (30) as previously described (4). The HYDRONMR computed diffusion tensor parameters and E-SL1m+Mg^{2+} structure were then used to compute ^{15}N R_2/R_1 values using in-house software.

Electrostatic calculations on NMR structures of the internal loop containing SL1m (1N8X (2), 2GM0 (21), and 2D17(22)) were performed using the non-linear Poisson Boltzmann equation to calculate the potential using the Delphi (Accelrys) module of Insight II (Molecular Simulations, Inc). The Debye-Huckel equation and full Coulombic boundary conditions were used to approximate the boundary potentials. Partial charges were obtained from the AMBER force field (58). Input parameters for the calculations were as described previously (59). A monovalent salt concentration of 0.025 M was assumed. The three-dimensional molecular surfaces were displayed using GRASP, provided by the Honig lab (http://honiglab.cpmc.columbia.edu) (Figure 2-10) (60).

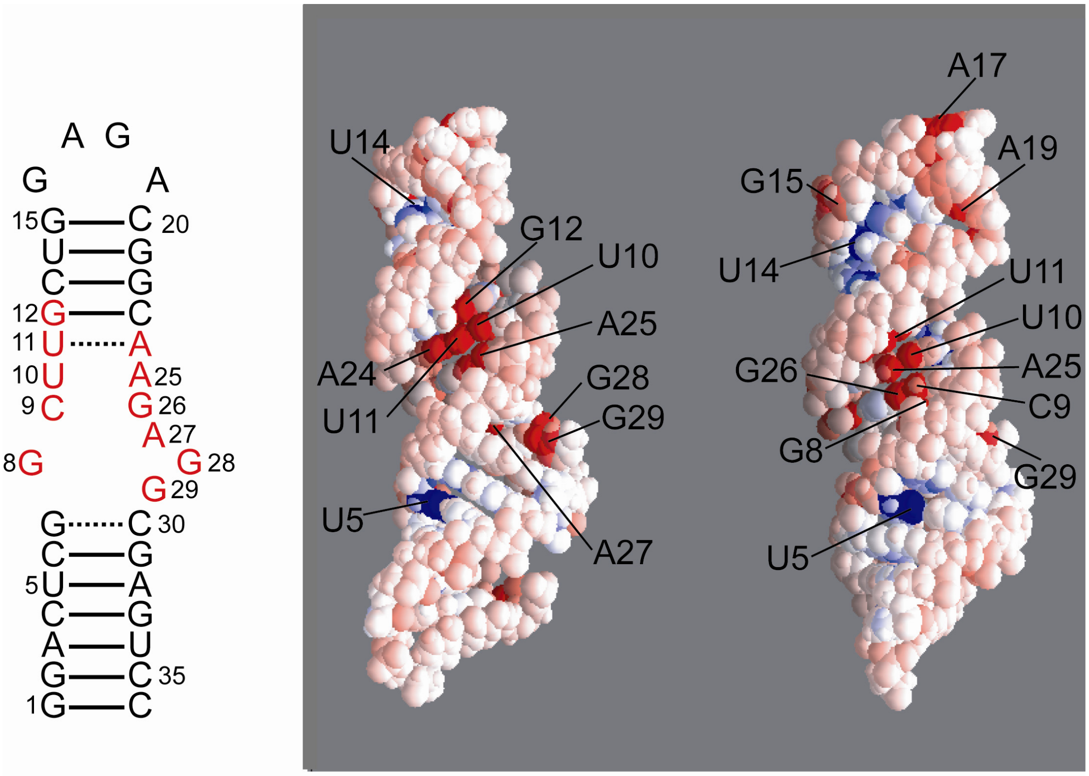

Figure 2-10. Electrostatic surface potential map of SL1m. It is computed using the previous GAGA-SL1m NMR structure (2) with negative and positive potentials indicated in red and blue, respectively. The negative electrostatic potential is particularly pronounced at residues that experience large Mg^{2+} induced chemical shift perturbations which are highlighted in red on the SL1m secondary structure.

This work has been published in Nucleic Acids Research (61)

2.5 References

1. Greatorex, J., Gallego, J., Varani, G. and Lever, A. (2002) Structure and stability of wild-type and mutant RNA internal loops from the SL-1 domain of the HIV-1 packaging signal. *J Mol Biol*, **322**, 543 - 557.
2. Lawrence, D.C., Stover, C.C., Noznitsky, J., Wu, Z. and Summers, M.F. (2003) Structure of the intact stem and bulge of HIV-1 Psi-RNA stem-loop SL1. *J Mol Biol*, **326**, 529-542.
3. Yuan, Y., Kerwood, D.J., Paoletti, A.C., Shubsda, M.F. and Borer, P.N. (2003) Stem of SL1 RNA in HIV-1: structure and nucleocapsid protein binding for a 1 x 3 internal loop. *Biochemistry*, **42**, 5259-5269.
4. Zhang, Q., Sun, X., Watt, E.D. and Al-Hashimi, H.M. (2006) Resolving the motional modes that code for RNA adaptation. *Science*, **311**, 653-656.
5. Dingley, A.J. and Grzesiek, S. (1998) Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide (2)J(NN) couplings. *Journal of the American Chemical Society*, **120**, 8293-8297.
6. Pervushin, K., Ono, A., Fernandez, C., Szyperski, T., Kainosho, M. and Wuthrich, K. (1998) NMR scaler couplings across Watson-Crick base pair hydrogen bonds in DNA observed by transverse relaxation optimized spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 14147-14151.
7. Hansen, M.R., Mueller, L. and Pardi, A. (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Structural Biology*, **5**, 1065-1074.
8. Clore, G.M., Starich, M.R. and Gronenborn, A.M. (1998) Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod- shaped viruses. *Journal of the American Chemical Society*, **120**, 10571-10572.
9. Losonczi, J.A., Andrec, M., Fischer, M.W.F. and Prestegard, J.H. (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. *Journal of Magnetic Resonance*, **138**, 334-342.
10. Musselman, C., Pitt, S.W., Gulati, K., Foster, L.L., Andricioaei, I. and Al-Hashimi, H.M. (2006) Impact of static and dynamic A-form heterogeneity on the determination of RNA global structural dynamics using NMR residual dipolar couplings. *J Biomol NMR*, **36**, 235-249.
11. Saupe, A. (1968) Recent results in the field of liquid crystals. *Angew. Chem., Int. Ed. Engl.*, **7**, 97-112.
12. Tolman, J.R., Al-Hashimi, H.M., Kay, L.E. and Prestegard, J.H. (2001) Structural and dynamic analysis of residual dipolar coupling data for proteins. *Journal of the American Chemical Society*, **123**, 1416-1424.
13. Al-Hashimi, H.M., Gosser, Y., Gorin, A., Hu, W., Majumdar, A. and Patel, D.J. (2002) Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. *J Mol Biol*, **315**, 95-102.

14. Al-Hashimi, H.M., Pitt, S.W., Majumdar, A., Xu, W. and Patel, D.J. (2003) Mg2+-induced variations in the conformation and dynamics of HIV-1 TAR RNA probed using NMR residual dipolar couplings. *J Mol Biol*, **329**, 867-873.
15. Pitt, S.W., Majumdar, A., Serganov, A., Patel, D.J. and Al-Hashimi, H.M. (2004) Argininamide binding arrests global motions in HIV-1 TAR RNA: comparison with Mg2+-induced conformational stabilization. *J Mol Biol*, **338**, 7-16.
16. Pitt, S.W., Zhang, Q., Patel, D.J. and Al-Hashimi, H.M. (2005) Evidence that electrostatic interactions dictate the ligand-induced arrest of RNA global flexibility. *Angew Chem Int Ed Engl*, **44**, 3412-3415.
17. Getz, M.M., Andrews, A.J., Fierke, C.A. and Al-Hashimi, H.M. (2006) Structural plasticity and Mg2+ binding properties of RNase P P4 from combined analysis of NMR residual dipolar couplings and motionally decoupled spin relaxation. *Rna*.
18. Sibille, N., Pardi, A., Simorre, J.P. and Blackledge, M. (2001) Refinement of local and long-range structural order in theophylline-binding RNA using C-13-H-1 residual dipolar couplings and restrained molecular dynamics. *Journal of the American Chemical Society*, **123**, 12135-12146.
19. Mollova, E.T., Hansen, M.R. and Pardi, A. (2000) Global structure of RNA determined with residual dipolar couplings. *Journal of the American Chemical Society*, **122**, 11561-11562.
20. Tolman, J.R. and Al-Hashimi, H.M. (2003) In Webb, G. A. (ed.), *Annual Reports on NMR Spectroscopy*. Academic Press, Vol. 51, pp. 105-166.
21. Ulyanov, N.B., Mujeeb, A., Du, Z., Tonelli, M., Parslow, T.G. and James, T.L. (2006) NMR structure of the full-length linear dimer of stem-loop-1 RNA in the HIV-1 dimer initiation site. *J Biol Chem*, **281**, 16168-16177.
22. Baba, S., Takahashi, K., Noguchi, S., Takaku, H., Koyanagi, Y., Yamamoto, N. and Kawai, G. (2005) Solution RNA structures of the HIV-1 dimerization initiation site in the kissing-loop and extended-duplex dimers. *J Biochem*, **138**, 583-592.
23. Clever, J.L. and Parslow, T.G. (1997) Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J Virol*, **71**, 3407-3414.
24. Clever, J.L., Wong, M.L. and Parslow, T.G. (1996) Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA. *J Virol*, **70**, 5902-5908.
25. Clever, J., Sassetti, C. and Parslow, T.G. (1995) RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *J Virol*, **69**, 2101 - 2109.
26. Tolman, J.R., Flanagan, J.M., Kennedy, M.A. and Prestegard, J.H. (1997) NMR evidence for slow collective motions in cyanometmyoglobin. *Nature Structural Biology*, **4**, 292-297.
27. Zhang, Q., Throolin, R., Pitt, S.W., Serganov, A. and Al-Hashimi, H.M. (2003) Probing motions between equivalent RNA domains using magnetic field induced residual dipolar couplings: accounting for correlations between motions and alignment. *J Am Chem Soc*, **125**, 10530-10531.
28. Clore, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C. and Gronenborn, A.M. (1990) Deviations from the Simple Two-Parameter Model-Free Approach to the

Interpretation of Nitrogen-15 Nuclear Magnetic Relaxation of Proteins. *Journal of the American Chemical Society*, **112**, 4989-4991.

29. Lipari, G. and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *Journal of the American Chemical Society*, **104**, 4546-4559.
30. Garcia de la Torre, J., Huertas, M.L. and Carrasco, B. (2000) HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J Magn Reson*, **147**, 138-146.
31. Zapp, M.L., Stern, S. and Green, M.R. (1993) Small molecules that selectively block RNA binding of HIV-1 Rev protein inhibit Rev function and viral production. *Cell*, **74**, 969-978.
32. McPike, M.P., Goodisman, J. and Dabrowiak, J.C. (2002) Footprinting and circular dichroism studies on paromomycin binding to the packaging region of human immunodeficiency virus type-1. *Bioorg Med Chem*, **10**, 3663-3672.
33. McPike, M.P., Sullivan, J.M., Goodisman, J. and Dabrowiak, J.C. (2002) Footprinting, circular dichroism and UV melting studies on neomycin B binding to the packaging region of human immunodeficiency virus type-1 RNA. *Nucleic Acids Res*, **30**, 2825-2831.
34. Takahashi, K.I., Baba, S., Chattopadhyay, P., Koyanagi, Y., Yamamoto, N., Takaku, H. and Kawai, G. (2000) Structural requirement for the two-step dimerization of human immunodeficiency virus type 1 genome. *Rna*, **6**, 96-102.
35. Aci, S., Mazier, S. and Genest, D. (2005) Conformational pathway for the kissing complex → extended dimer transition of the SL1 stem-loop from genomic HIV-1 RNA as monitored by targeted molecular dynamics techniques. *J Mol Biol*, **351**, 520-530.
36. Theilleux-Delalande, V., Girard, F., Huynh-Dinh, T., Lancelot, G. and Paoletti, J. (2000) The HIV-1(Lai) RNA dimerization. Thermodynamic parameters associated with the transition from the kissing complex to the extended dimer. *Eur J Biochem*, **267**, 2711-2719.
37. Rist, M.J. and Marino, J.P. (2002) Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1. *Biochemistry*, **41**, 14762-14770.
38. Windbichler, N., Werner, M. and Schroeder, R. (2003) Kissing complex-mediated dimerisation of HIV-1 RNA: coupling extended duplex formation to ribozyme cleavage. *Nucleic Acids Res*, **31**, 6419-6427.
39. Amarasinghe, G.K., De Guzman, R.N., Turner, R.B. and Summers, M.F. (2000) NMR structure of stem-loop SL2 of the HIV-1 psi RNA packaging signal reveals a novel A-U-A base-triple platform. *J Mol Biol*, **299**, 145-156.
40. De Guzman, R.N., Wu, Z.R., Stalling, C.C., Pappalardo, L., Borer, P.N. and Summers, M.F. (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science*, **279**, 384 - 388.
41. Damgaard, C.K., Dyhr-Mikkelsen, H. and Kjems, J. (1998) Mapping the RNA binding sites for human immunodeficiency virus type-1 gag and NC proteins within the complete HIV-1 and -2 untranslated leader regions. *Nucleic Acids Res*, **26**, 3667-3676.

42. Mirambeau, G., Lyonnais, S., Coulaud, D., Hameau, L., Lafosse, S., Jeusset, J., Justome, A., Delain, E., Gorelick, R.J. and Le Cam, E. (2006) Transmission electron microscopy reveals an optimal HIV-1 nucleocapsid aggregation with single-stranded nucleic acids and the mature HIV-1 nucleocapsid protein. *J Mol Biol*, **364**, 496-511.
43. Takahashi, K., Baba, S., Koyanagi, Y., Yamamoto, N., Takaku, H. and Kawai, G. (2001) Two basic regions of NCp7 are sufficient for conformational conversion of HIV-1 dimerization initiation site from kissing-loop dimer to extended-duplex dimer. *J Biol Chem*, **276**, 31274-31278.
44. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) Nmrpipe - a Multidimensional Spectral Processing System Based On Unix Pipes. *Journal of Biomolecular Nmr*, **6**, 277-293.
45. Johnson, B.A. and Blevins, R.A. (1994) Nmr View - A Computer-Program For The Visualization And Analysis Of Nmr Data. *Journal Of Biomolecular Nmr*, **4**, 603-614.
46. Goddard, T.D. and Kneller, D.G. *SPARKY 3*. University of California, San Franscisco.
47. Furtig, B., Richter, C., Wohnert, J. and Schwalbe, H. (2003) NMR spectroscopy of RNA. *Chembiochem*, **4**, 936-962.
48. Meissner, A., Duus, J.O. and Sorensen, O.W. (1997) Spin-state-selective excitation. Application for E.COSY-type measurement of J(HH) coupling constants. *Journal of Magnetic Resonance*, **128**, 92-97.
49. Meissner, A. and Sorensen, O.W. (1999) The role of coherence transfer efficiency in design of TROSY- type multidimensional NMR experiments. *Journal of Magnetic Resonance*, **139**, 439-442.
50. Yip, G.N. and Zuiderweg, E.R. (2004) A phase cycle scheme that significantly suppresses offset-dependent artifacts in the R2-CPMG 15N relaxation experiment. *J Magn Reson*, **171**, 25-36.
51. Yip, G.N. and Zuiderweg, E.R. (2005) Improvement of duty-cycle heating compensation in NMR spin relaxation experiments. *J Magn Reson*.
52. Saupe, A. and Englert, G. (1963) High-resolution nuclear magnetic resonance spectra of oriented molecules. *Physical Review Letters*, **11**, 462-464.
53. Al-Hashimi, H.M., Valafar, H., Terrell, M., Zartler, E.R., Eidsness, M.K. and Prestegard, J.H. (2000) Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings. *Journal of Magnetic Resonance*, **143**, 402-406.
54. Hansen, A.L. and Al-Hashimi, H.M. (2006) Insight into the CSA tensors of nucleobase carbons in RNA polynucleotides from solution measurements of residual CSA: towards new long-range orientational constraints. *J Magn Reson*, **179**, 299-307.
55. Zweckstetter, M. and Bax, A. (2000) Predicition of sterically induced alignment in a dilute liquid crystalline phase; aid to protein structure determination by NMR. *J . Am. Chem. Soc.*, **122**, 3791-3792.
56. Mandel, A.M., Akke, M. and Palmer, A.G. (1995) Backbone Dynamics of Escherichia-Coli Ribonuclease Hi - Correlations With Structure and Function in an Active Enzyme. *Journal of Molecular Biology*, **246**, 144-163.

57. Gonzalez, R.L.J. and Tinoco, I.J. (1999) Solution structure and thermodynamics of a divalent metal Ion binding site in an RNA pseudoknot. *J. Mol. Biol.*, **289**, 1267-1282.
58. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) A 2nd Generation Force-Field For The Simulation Of Proteins, Nucleic-Acids, And Organic-Molecules. *Journal Of The American Chemical Society*, **117**, 5179-5197.
59. Chin, K., Sharp, K.A., Honig, B. and Pyle, A.M. (1999) Calculating the electrostatic properties of RNA provides new insights into molecular interactions and function. *Nat Struct Biol*, **6**, 1055-1061.
60. Nicholls, A., Sharp, K. and Honig, B. (1991) Protein folding and association:insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281-296.
61. Sun, X., Zhang, Q. and Al-Hashimi, H.M. (2007) Resolving fast and slow motions in the internal loop containing stem-loop 1 of HIV-1 that are modulated by Mg2+ binding: role in the kissing-duplex structural transition. *Nucleic Acids Res*, **35**, 1698-1713.

Chapter 3

Structural dynamics of SL1 kissing and duplex dimers in the absence and presence of Mg^{2+}

3.1 Introduction

In our previous study, we substituted the wild-type apical loop of SL1 with a stable GAGA tetraloop that serves to prevent spontaneous dimer formation. Our results revealed that two types of motions are induced by the internal loop of SL1; (1) micro-to-millisecond intra-helix motions that lead to partial melting of base-pairs in the upper stem and (2) nanosecond inter-domain motions which may help bring strands into spatial register during strand exchange. Furthermore, we showed that Mg^{2+} specifically binds to the internal loop and significantly quenches both types of motion. Though chemical shift mapping data indicates that the apical loop replacement does not impact the integrity of the stem-internal loop-stem element in SL1, it is important that we verify that the dynamical and Mg^{2+} binding characteristics are preserved in the kissing complex. In addition, our studies so far provided no information regarding the key structural dynamics and Mg^{2+} binding properties of the apical loop and how this might differ in the kissing and duplex dimers. As described in the introduction, there has been a great deal of discrepancy regarding the conformation of the kissing dimer and its comparison with the duplex dimer. Very little studies have addressed and compared the dynamical properties of the two dimers which may be at the heart of the observed discrepancies. Intrinsic dynamics in the kissing dimer may play a role in the transition to the duplex whereas the high stability of a duplex dimer may be important for a mature HIV-1 particle.

In this chapter, we investigated the structural dynamics of SL1 in the native context of kissing and duplex dimers and examined the impact of Mg^{2+} binding. These quantitative studies of dynamics are made challenging by the more than two-fold increase in the SL1 size. However, these studies are now greatly facilitated by our initial characterization of the SL1m monomer.

3.2 Results

3.2.1 Kissing and duplex dimers share similar structural dynamics and Mg^{2+} binding properties in the internal but not apical loop

To examine possible differences between the kissing and duplex SL1 dimers, we first compared 2D HSQC spectra of the two samples prepared as described in the methods section. The majority of resonances in the helices and internal loop could be assigned by overlaying spectra of the dimer with those of the assigned SL1m mutant. For the apical loop, we assigned those resonances based on previous published SL1 dimers studies (1-3). As shown in figure 3-1, we observed significantly greater overlap in spectra of the dimers as compared to SL1m (Chapter 2).

As described previously in Chapter 2, the common resonances from the kissing SL1 dimer (excluding those near the apical loop) overlay very well with those of SL1m both in the presence and absence of Mg^{2+}. As shown in figure 3-1A, the far majority of these resonances in the stem and internal loop are also similar when comparing kissing and duplex dimers both in the absence and presence of 8mM Mg^{2+}. However, notable chemical shift differences are observed for all residues in the apical loop and a number of neighboring residues, such as G15 and G21. For example, while the C8H8 resonance of the loop residue LA2 has an unusual chemical shift in the kissing dimer, which has been ascribed to protonation of the N1 base nitrogen of 5' adenine in the apical loop (3) (Figure 3-1), this was not the case for the duplex dimer. The differences observed for G21 might be due to different local structures of the apical loop and adjacent stem region in both conformations (4-6). As we discuss in more detail below, there are marked differences in the broadening pattern for resonances belonging to apical and neighboring residues in the kissing and duplex dimer that suggest differences in structural dynamics. For example, in the kissing dimer, both 5' As and neighboring G15 are clearly exchange broadened consistent with μs-ms motions, but this is not the case in the duplex dimer. The motions likely reflect the looping in and out of the 5' As as observed/proposed in previous MD and structural studies (1,7-11).

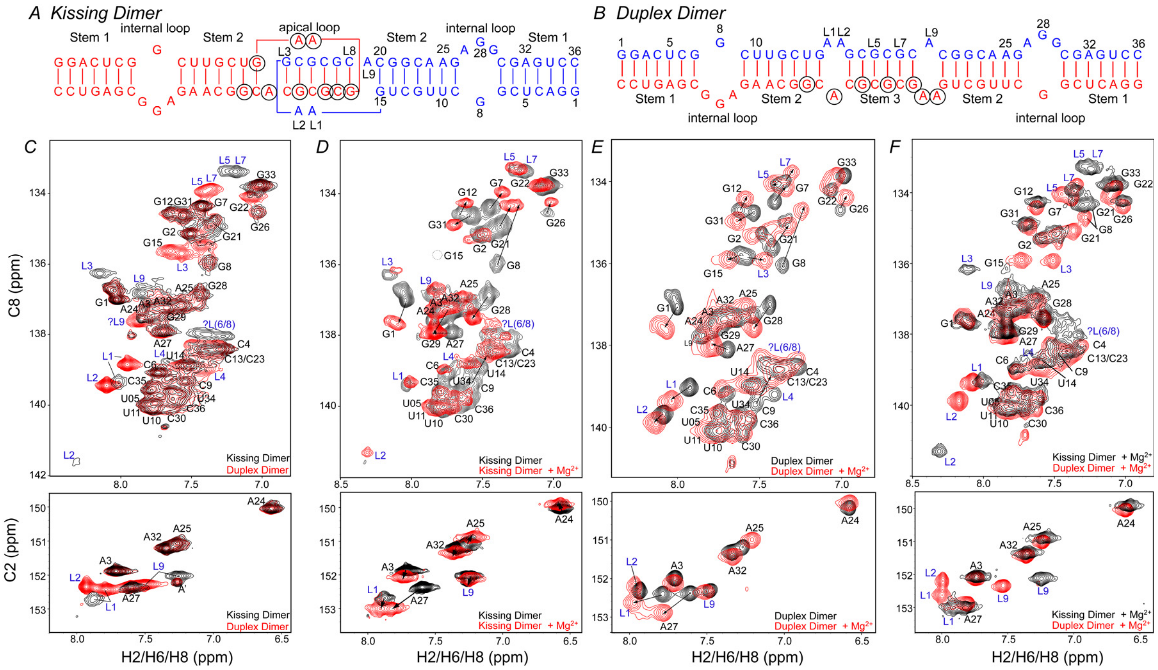

Figure 3-1. NMR spectra of SL1 kissing and duplex dimers in the absence and presence of Mg^{2+}. (A-B) Secondary structures of kissing (A) and duplex (B) SL1 dimers. (C) Chemical shift mapping of the kissing and duplex dimers. Shown are overlays of 2D ^1H-^{13}C spectra of kissing (in black) and duplex (in red) dimers. (D-E) Chemical shift mapping of Mg^{2+} binding to the kissing (D) and duplex dimers (E). Shown are overlays of 2D ^1H-^{13}C spectra in the absence (in black) and presence (in red) of 8mM Mg^{2+}.(F) Chemical shift mapping of the kissing and duplex dimers in the presence of 8mM Mg^{2+}. Shown are overlays of 2D ^1H-^{13}C spectra of kissing (in black) and duplex (in red) dimers.

The kissing and duplex dimers also exhibited marked differences in the Mg^{2+} (8 mM) induced chemical shift perturbations in the apical loop and neighboring residues. Upon Mg^{2+} association, relatively small chemical shift perturbations were observed in the apical loop of the kissing dimer. In contrast, large chemical shift perturbations were observed at the apical loop in the duplex dimer, especially for the unpaired As. Together, these data indicate that the apical loop in kissing and duplex dimer differs in their conformational dynamics and Mg^{2+} binding properties, but that they otherwise share very similar properties for internal loop and stem.

3.2.2 Temperature dependence of kissing and duplex dimer structural dynamics in the absence and presence of Mg^{2+}

To further understand the differences in the apical loop between the kissing and duplex dimers, we examined the chemical shift perturbations induced by changing the temperature. This led us to also observe marked differences in the exchange broadening behavior at the apical loop and neighboring residues indicating differences in conformational dynamics. Overall, the temperature dependence of the internal loop agreed very well with our studies of SL1m for both the kissing and duplex dimers. Apical loop residues LA1 and LA2 experienced intermediate or slow exchange in the kissing dimer but not in the duplex dimer. The exchange broadening in these loops in the kissing dimer are not markedly reduced with temperature. Rather, we observed a second set of peaks for these residues possibly denoting a second state that is more accessible at higher temperature. On the other hand, in the duplex dimer, these unpaired As do not undergo significant changes in intensity but instead experience large chemical shift perturbations, consistent with rapid exchange between two or more states (Figure 3-2).

The addition of 8 mM Mg^{2+} significantly reduced the exchange broadening observed in the apical loop of the kissing dimer but did not lead to any significant chemical perturbations. This indicates that Mg^{2+} reduces conformational exchange by stabilizing the apical loop conformation and the overall kissing dimer. Interestingly, in the presence of Mg^{2+}, we did not observe the second set of peaks for residues LA1 and LA2, again indicating that Mg^{2+} stabilizes a specific apical loop conformation. In contrast, LA1 and LA2 in the duplex dimer have different response to Mg^{2+}. At low temperature, LA1 and LA2 in the duplex dimer showed pronounced exchange broadening in the presence of Mg^{2+}, which were not observed in the absence of Mg^{2+} or in the kissing dimer. This exchange broadening was significantly reduced with increasing temperature. The exchange broadening found at low temperature might be due to the reduced exchange rate between two states of LA1 and LA2 or differences in Mg^{2+} binding.

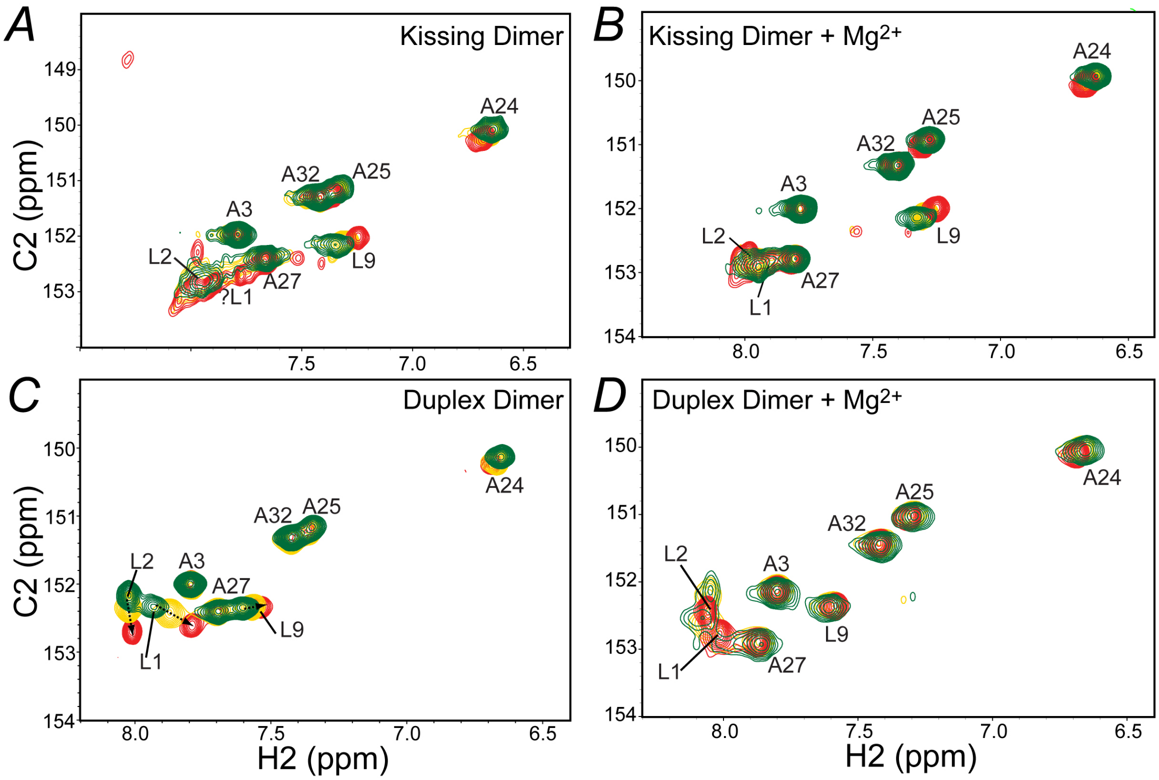

Figure 3-2. Temperature dependent of kissing and duplex dimers in the presence and absence of Mg^{2+}. Unpaired As from the apical loop region of kissing and duplex dimers presences different structural dynamics in the absence and presence of Mg^{2+} at various temperatures, 15°C (in green), 25°C (in yellow), and 37°C (in red). Representative spectra of C2-H2 regions in kissing (A-B) and duplex (C-D) dimers.

To gain further insight into the dynamics of the kissing and duplex dimers, we quantitatively measured resonance intensities at 25°C in the absence and presence of Mg^{2+} and normalized their values as described in chapter 2 (Figure 3-3). As noted above, exchange broadening is observed for many residues in the apical loop in the kissing dimer which is reduced upon Mg^{2+} binding. Less broadening is observed in the duplex dimer and in fact we observed elevated intensities for C2H2 in LA1 and LA2 that are consistent with fast ps-ns motions. Interestingly, these intensities were reduced upon Mg^{2+} binding, but the intensities of the corresponding C8H8 resonance remained relatively high.

In both the kissing and duplex dimer, high intensities are observed for stem I as compared to stem II but this difference is reduced upon Mg^{2+} binding. The latter may be attributed to inter-domain motions observed in SL1m across the internal loop in the absence of Mg^{2+}. Here, the two stem II helices effectively act as an "elongated" reference helix. Unfortunately, independent confirmation of dynamical averaging in the base of A27 and G29 was not possible due to spectral overlap in the non-elongated uniformly labeled SL1 samples.

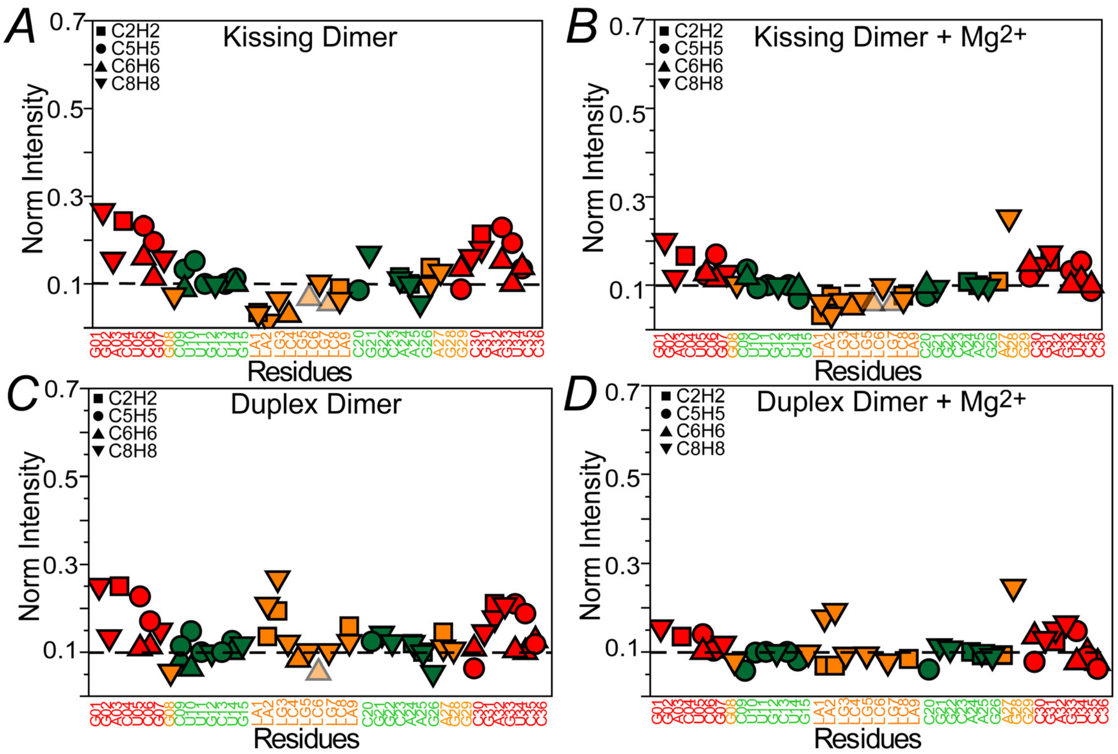

Figure 3-3. Local and domain motions in kissing and duplex dimers in the absence and presence of Mg^{2+}. Shown are normalized resonance intensities as function of residue in

the absence and presence of 8 mM Mg^{2+} for kissing dimer (A-B) and duplex dimer (C-D) at 25°C. Resonance intensities are obtained from non-constant time ^{1}H-^{13}C HSQC experiments. The intensity for a given type of C-H vector is normalized relative to a minimum value of 0.1 indicated by a horizontal line. Peaks that do not exhibit exchange broadening as inferred from the temperature-induced perturbations were not used in the normalization.

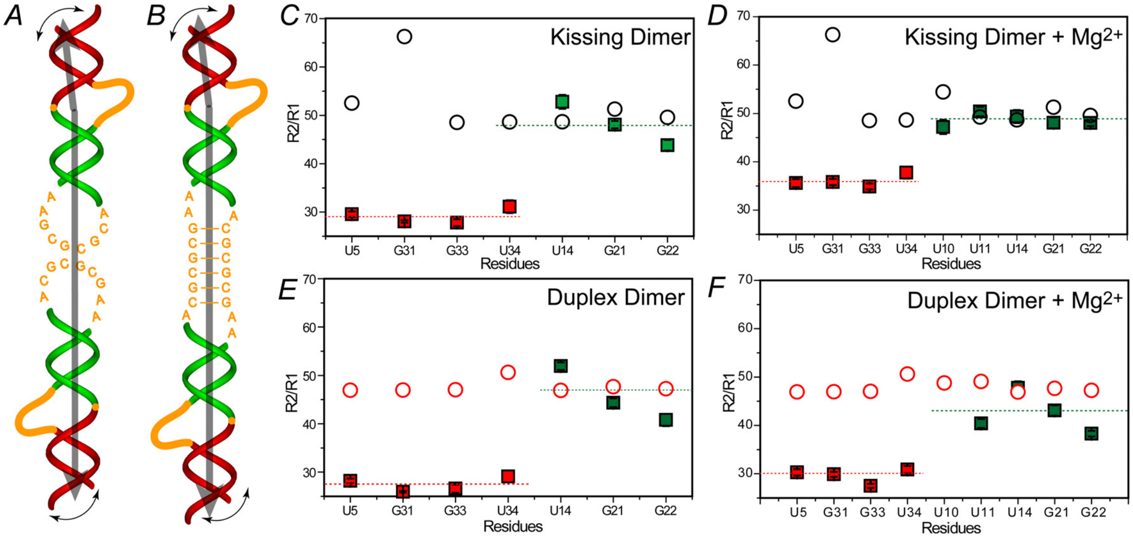

Figure 3-4. ^{15}N relaxation data measured in kissing and duplex dimers in absence(12) and presence of Mg^{2+}. (A-B) Secondary structures of metastable kissing (A) and stable duplex (B) SL1 dimers. (C-D) R_2/R_1 values for kissing dimer in the absence and presence of 8mM Mg^{2+}. (E-F) R_2/R_1 values for duplex dimer in the absence and presence of 8mM Mg^{2+}. Values predicted using hydrodynamic calculations are shown in black (kissing dimer) and red (duplex dimer) circles. Average values of R_2/R_1 from stem1 and stem2 are shown in red and green dash lines.

3.2.3 Inter-domain motions by ^{15}N spin relaxation

We used ^{15}N spin relaxation measurements to further examine local and inter-domain motions in the SL1 kissing and duplex dimers. Rather than elongating the samples, which would lead to RNA of unmanageable size, we exploit the inherent asymmetry of the dimers, in which the pair of stem II helices are effectively twice as long as those of stem I, providing a basis for decoupling inter-domain motions from overall rotational diffusion (Figure 3-4A and B).

In our previous study of dynamics of SL1, the ^{15}N relaxation data measured in E-SL1m revealed nanosecond inter-stem motions that were arrested by Mg^{2+}. Similarly, in both dimer conformations, the stem I R_2/R_1 ratios were significantly smaller than values measured in stem II or values computed hydrodynamically (13) when assuming a static structure. Thus, the ^{15}N relaxation data measured in kissing and duplex dimers confirmed that the inter-domain motions induced by the internal loop in SL1m are also present in the dimers. In both cases, the addition of Mg^{2+} led to a reduction in the differences observed in stems I and II consistent with a reduction in inter-domain motions, as we reported for SL1m.

3.3 Discussion

The highly conserved SL1 domain of the HIV-1 RNA genome plays important roles to form the dimeric virus RNA, which includes initiating the loop-loop kissing via complementary loop sequence and converting it into the thermodynamically stable duplex dimer. The GC rich apical loop is critical for kissing dimer formation and the flanking unpaired adenines are believed to be critical for the stable loop-loop interaction. We previously revealed that the highly conserved G-AGG internal loop induces internal motions, including melting dynamics of the upper stem and inter-domain motions that may be important for driving the kissing to duplex structural transition. By studying wild type SL1, our results here confirm that similar internal motions are induced by the internal loop in the context of both kissing and duplex dimers. This result establishes that the internal loop is structurally, dynamically, and functionally decoupled from the other key site, the apical loop in SL1.

There has been a great deal of discrepancy regarding the conformation of the kissing dimer and specifically the relative orientation of the upper helical stems and local conformation of the apical loop adenines which have been shown to adopt either a looped in or out conformation (1,4,10,14). A very likely source for these discrepancies is that the apical loop is highly flexible and depending on conditions such as salt concentration, presence or absence of divalent ions, temperature, different conformations can be favored by solution NMR or X-ray crystallography. Yet despite this possibility, very few experimental studies have quantitatively examined the dynamics of the kissing apical loop and the impact of Mg^{2+} binding.

Our results clearly show that the apical loop, and particularly the 5' As are exchange broadened indicating that they experience a slow μs-ms dynamical process (Figure 3-1). In principle, this could correspond to slow exchange between dimer and monomer SL1 and the broadening would thus reflect dissociation of the loop-loop interaction. Indeed, the kissing dimer runs on native PAGE in the absence of Mg^{2+} as a combination of monomer and dimer. If the kissing dimer exists in large fraction as a monomer under the NMR conditions, one can expect that the observed overall correlation time and thus the ^{15}N R_2/R_1 values for the kissing dimer should be significantly smaller than that observed in the duplex dimer. This is however not observed indicating that any monomer population can be expected to be rather small (<10%) (Figure 3-4). While exchange involving a small fraction of a monomer species could account for the observed broadening, the broadening in the kissing dimer could also reflect local "melting" dynamics at the apical loop in addition to complete dissociation of the two monomers. Evidence for this sort of dynamics comes from the observation of a degree of exchange broadening in the duplex dimer and persistence of broadening in the kissing dimer even after addition of Mg^{2+} which is known to stabilize the kissing dimer (see below) (Figure 3-3).

It is very likely that the adenine residues adopt different conformations as the various G-C loop-loop base-pairs disruptions. In the intact kissing dimer conformation, As are likely to resemble the various X-ray structures and adopt a looped out conformation (10,14,15),whereas towards the monomeric state they are more likely to adopt a looped in conformation as observed in the NMR studies (2,4). In the former case, the stem II helices are expected to be coaxially stacked whereas in the latter case, there is room for inter-helical bending. This gives rise to a coupling between the status of the As and inter-helical bending. Indeed, we generally observe that bent structures arise when those As stably stayed looped in and linear structures when they are looped out (4) or toward looping out state(2,10,14,15). It should be noted also that NOEs for the looped in state can be detected even if sampled transiently whereas there are likely fewer characteristic NOEs that define the looped out state.

Mg^{2+} is known to stabilize the metastable kissing dimer, though the effects are more pronounced for sub-type A compared to the sub-type B used in our studies (16). Consistent with this observation, the kissing dimer runs more cleanly as a dimer under native PAGE in the presence of Mg^{2+} and we observed less exchange broadening in the kissing dimer following addition of 8 mM Mg^{2+} (Figure 3-3). However, some exchange broadening persists even in the presence of Mg^{2+} consistent with a local dynamical process (Figure 3-3). Likewise, Mg^{2+} did not lead to a significant increase in the ^{15}N R_2/R_1 values measured in the kissing dimer (Figure 3-4), again consistent with the idea that any monomer population is very small.

The idea that the observed broadening in the kissing dimer arises from dissociation of the loop-loop interaction is also supported by the significant reduction in exchange broadening observed in the duplex dimer which runs as a single band in PAGE. In the duplex dimer, the two 5' As undergo extensive local motions at the ps-ns timescales (Figure 3-1), consistent with a looped out conformation that has been observed in several structures of the duplex dimer (17).

Several molecular dynamics simulation studies have examined the dynamics of the kissing and duplex dimer and their results are in good agreement with what we detect by NMR. For example, all MD studies have shown that the NMR structure appears more stable than the crystal structure, and those bulged-out unpaired As in the crystal structure switch towards a bulge-in conformation as observed in the NMR solved structure (7,9,18-20).

The dynamics at the As provides a flexible hinge for modulating the relative orientation of the two stem II helices that are connected via the loop-loop interaction in the kissing dimer. This provides support for our model for structural isomerization (Figure 2-9 Chapter 2) in which individual monomeric units come into close proximity for strand exchange by rotating along the C2 symmetry axis without disrupting the kissing loop. On the other hand, in the duplex dimer conformation, the six GC residues form a stable helix.

In conclusion, our results expose important differences in the structural, dynamical, and Mg^{2+} binding properties of the kissing and duplex dimer that have not been seen before, explain existing discrepancies as well as lend support to our dynamical model for the kissing to duplex transition.

Table 3-1: ^{15}N relaxation data measured in kissing dimer in the absence and presence of 8mM Mg^{2+}

Kissing Dimer			Kissing Dimer + Mg^{2+}		
Residue	R2 (Hz)	R1 (Hz)	Residue	R2 (Hz)	R1 (Hz)
U5 (N3)	23.07 ± 0.23	0.78 ± 0.01	U5 (N3)	26.49 ± 0.23	0.74 ± 0.02
G31 (N1)	24.05 ± 0.11	0.86 ± 0.01	G31 (N1)	28.15 ± 0.35	0.79 ± 0.01
G33 (N1)	24.37 ± 0.50	0.88 ± 0.01	G33 (N1)	28.38 ± 0.27	0.81 ± 0.01
U34 (N3)	22.96 ± 0.85	0.74 ± 0.01	U34 (N3)	28.01 ± 0.50	0.74 ± 0.02
U10 (N3)	N/A	N/A	U10 (N3)	30.297 ± 0.38	0.64 ± 0.02
U11 (N3)	N/A	N/A	U11 (N3)	31.16 ± 0.35	0.62 ± 0.01
U14 (N3)	32.30 ± 0.21	0.61 ± 0.02	U14 (N3)	32.95 ± 0.24	0.67 ± 0.02
G21 (N1)	32.72 ± 0.23	0.68 ± 0.01	G21 (N1)	33.70 ± 0.40	0.70 ± 0.01
G22 (N1)	29.30 ± 0.73	0.67 ± 0.003	G22 (N1)	31.45 ± 0.29	0.65 ± 0.01

Table 3-2: ^{15}N relaxation data measured in duplex dimer in the absence and presence of 8mM Mg^{2+}

Duplex Dimer			Duplex Dimer + Mg^{2+}		
Residue	R2 (Hz)	R1 (Hz)	Residue	R2 (Hz)	R1 (Hz)
U5 (N3)	23.00 ± 0.26	0.82 ± 0.01	U5 (N3)	24.87 ± 0.17	0.82 ± 0.02
G31 (N1)	23.65 ± 0.09	0.91 ± 0.003	G31 (N1)	26.37 ± 0.40	0.88 ± 0.01
G33 (N1)	24.21 ± 0.75	0.911 ± 0.01	G33 (N1)	25.24 ± 0.31	0.92 ± 0.01
U11 (N3)	N/A	N/A	U11 (N3)	24.59 ± 0.58	0.70 ± 0.01
U34 (N3)	22.73 ± 0.91	0.78 ± 0.01	U34 (N3)	28.19 ± 0.36	0.80 ± 0.02
U14 (N3)	32.34 ± 0.28	0.62 ± 0.01	U14 (N3)	33.439 ± 0.38	0.70 ± 0.01
G21 (N1)	32.57 ± 0.16	0.73 ± 0.01	G21 (N1)	34.18 ± 0.37	0.79 ± 0.02
G22 (N1)	29.71 ± 0.90	0.73 ± 0.01	G22 (N1)	28.63 ± 0.35	0.75 ± 0.01

3.4 Materials and methods

3.4.1 Sample preparation

Uniformly $^{13}C/^{15}N$ labeled SL1 NMR samples (0.6-1.0 mM) were prepared using standard *in vitro* transcription reactions as previously described(12). The NMR buffer contained 15 mM sodium phosphate, 25 mM sodium chloride, 0.1 mM EDTA, and pH ~ 6.8 in 10% D_2O. The kissing dimer was prepared by incubating less than 0.1mM desalted RNA in water at 95ºC for 10 mins, followed by cooling on ice, and then concentrated and exchanged into ~0.8 mM in NMR buffer using centricon devices with ultracel YM-3 membrance (Millipore). The duplex dimer dimer was prepared by incubating ~0.8mM desalted RNA dissolved in water at 95ºC for 5 mins followed by slow cooling at room temperature.

3.4.2 NMR spectroscopy

All NMR experiments were preformed on an Avance Bruker 600 MHz NMR spectrometer equipped with a triple-resonance 5 mm cryogenic probe. Unless stated otherwise, all NMR experiments were conducted at 298 K. NMR spectra were processed using NMRPipe/NMRDraw, analyzed using NMRView, and overlaid using SPARKY 3. All non-exchangeable resonances were assigned using standard methods and found to be in excellent agreement with those reported previously for the kissing and duplex construct. We assigned the non-exchangeable resonances from the both stems and internal loop of SL1 kissing and duplex dimers based on our SL1 monomer assignments, since they have a very good agreement on the stem-bugle-stem region; we assigned the apical loop region of the kissing and duplex dimers based on the previous NMR studies (1-3,21).

Imino ^{15}N longitudinal (R_1) and transverse ($R_{2(\mathrm{CPMG})}$) relaxation rates and ^{1}H-^{15}N were measured in kissing and duplex dimers in the absence and presence of 8 mM Mg^{2+} using well established 2D experiments closely following the procedure described previously for free E-SL1m. For R_1 experiments, the relaxation delays were 0.06, 0.12, 0.24, 0.48, 0.64, 0.80, and 1.2 s. The $R_{2(\mathrm{CPMG})}$ experiment employed a recently described $[0013]^N$ phase cycle in the Carr-Purcell-Meiboom-Gill (CPMG) pulse train for suppressing artifacts arising from off-resonance effects. The relaxation delays were 0.0062, 0.0124, 0.0248, 0.0372, 0.0496, 0.0620, and 0.0744 s. The final R_2 values were corrected to account for off-resonance effects as previously described. Uncertainties were verified based on duplicate measurements of relaxation data. Mg^{2+} chemical shift titrations were performed by recording 2D ^{1}H-^{13}C or ^{1}H-^{15}N HSQC spectra following incremental increases in the $MgCl_2$ concentration (0.1, 0.5, 1.0, 2.0, 4.0 and 8.0 mM).

3.5 References

1. Kieken, F., Arnoult, E., Barbault, F., Paquet, F., Huynh-Dinh, T., Paoletti, J., Genest, D. and Lancelot, G. (2002) HIV-1(Lai) genomic RNA: combined used of NMR and molecular dynamics simulation for studying the structure and internal dynamics of a mutated SL1 hairpin. *Eur Biophys J*, **31**, 521-531.
2. Kieken, F., Paquet, F., Brule, F., Paoletti, J. and Lancelot, G. (2006) A new NMR solution structure of the SL1 HIV-1Lai loop-loop dimer. *Nucleic Acids Res*, **34**, 343-352.
3. Mihailescu, M.R. and Marino, J.P. (2004) A proton-coupled dynamic conformational switch in the HIV-1 dimerization initiation site kissing complex. *Proc Natl Acad Sci U S A*, **101**, 1189-1194.
4. Mujeeb, A., Clever, J.L., Billeci, T.M., James, T.L. and Parslow, T.G. (1998) Structure of the dimer initiation complex of HIV-1 genomic RNA. *Nat Struct Biol*, **5**, 432-436.
5. Girard, F., Barbault, F., Gouyette, C., Huynh-Dinh, T., Paoletti, J. and Lancelot, G. (1999) Dimer initiation sequence of HIV-1Lai genomic RNA: NMR solution structure of the extended duplex. *J Biomol Struct Dyn*, **16**, 1145-1157.
6. Mujeeb, A., Parslow, T.G., Zarrinpar, A., Das, C. and James, T.L. (1999) NMR structure of the mature dimer initiation complex of HIV-1 genomic RNA. *FEBS Lett*, **458**, 387-392.
7. Beaurain, F. and Laguerre, M. (2003) MD studies of the DIS/DIS kissing complex solution and x-ray structures. *Oligonucleotides*, **13**, 501-514.
8. La Penna, G., Genest, D. and Perico, A. (2003) Modeling the dynamics of the solvated SL1 domain of HIV-1 genomic RNA. *Biopolymers*, **69**, 1-14.
9. Sarzynska, J., Reblova, K., Sponer, J. and Kulinski, T. (2008) Conformational transitions of flanking purines in HIV-1 RNA dimerization initiation site kissing complexes studied by CHARMM explicit solvent molecular dynamics. *Biopolymers*, **89**, 732-746.
10. Ennifar, E. and Dumas, P. (2006) Polymorphism of bulged-out residues in HIV-1 RNA DIS kissing complex and structure comparison with solution studies. *J Mol Biol*, **356**, 771-782.
11. Ennifar, E., Paillart, J.C., Marquet, R., Ehresmann, B., Ehresmann, C., Dumas, P. and Walter, P. (2003) HIV-1 RNA dimerization initiation site is structurally similar to the ribosomal A site and binds aminoglycoside antibiotics. *J Biol Chem*, **278**, 2723-2730.
12. Zhang, Q., Sun, X., Watt, E.D. and Al-Hashimi, H.M. (2006) Resolving the motional modes that code for RNA adaptation. *Science*, **311**, 653-656.
13. Garcia de la Torre, J., Huertas, M.L. and Carrasco, B. (2000) HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J Magn Reson*, **147**, 138-146.
14. Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C. and Dumas, P. (2001) Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat Struct Biol*, **8**, 1064-1068.

15. Ennifar, E., Paillart, J.C., Bernacchi, S., Walter, P., Pale, P., Decout, J.L., Marquet, R. and Dumas, P. (2007) A structure-based approach for targeting the HIV-1 genomic RNA dimerization initiation site. *Biochimie*, **89**, 1195-1203.
16. Jossinet, F., Paillart, J.C., Westhof, E., Hermann, T., Skripkin, E., Lodmell, J.S., Ehresmann, C., Ehresmann, B. and Marquet, R. (1999) Dimerization of HIV-1 genomic RNA of subtypes A and B: RNA loop structure and magnesium binding. *Rna*, **5**, 1222-1234.
17. Ennifar, E., Yusupov, M., Walter, P., Marquet, R., Ehresmann, B., Ehresmann, C. and Dumas, P. (1999) The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. *Structure*, **7**, 1439-1449.
18. Aci, S., Mazier, S. and Genest, D. (2005) Conformational pathway for the kissing complex-->extended dimer transition of the SL1 stem-loop from genomic HIV-1 RNA as monitored by targeted molecular dynamics techniques. *J Mol Biol*, **351**, 520-530.
19. Aci, S., Gangneux, L., Paoletti, J. and Genest, D. (2004) On the stability of different experimental dimeric structures of the SL1 sequence from the genomic RNA of HIV-1 in solution: a molecular dynamics simulation and electrophoresis study. *Biopolymers*, **74**, 177-188.
20. Reblova, K., Fadrna, E., Sarzynska, J., Kulinski, T., Kulhanek, P., Ennifar, E., Koca, J. and Sponer, J. (2007) Conformations of flanking bases in HIV-1 RNA DIS kissing complexes studied by molecular dynamics. *Biophys J*, **93**, 3932-3949.
21. Lawrence, D.C., Stover, C.C., Noznitsky, J., Wu, Z. and Summers, M.F. (2003) Structure of the intact stem and bulge of HIV-1 Psi-RNA stem-loop SL1. *J Mol Biol*, **326**, 529-542.

Chapter 4

NC protein catalyzed isomerization of SL1 from kissing to duplex dimer in the presence of Mg^{2+}

4.1 Introduction

Our equilibrium studies on the SL1 kissing and duplex dimers along with other studies published in the literature led us to propose that the kissing to duplex transition occurs via a quadruplex intermediate which brings together two monomeric strands. Furthermore, we proposed that NC protein directly binds the internal loop which acts as the site for nucleating strand exchange. In this chapter, we examine the interaction between NC proteins with SL1 and use a time-resolved NMR strategy to monitor site-specifically and in real time the NC protein catalyzed kissing to duplex dimer transition. Our goal was to shed light onto the time course of the transition, including RNA-NC protein interactions, and obtains evidence for our proposed intermediate.

4.2 Results

4.2.1 NC protein binds the kissing SL1 internal loop in the presence of Mg^{2+}

We first used NMR to examine if NC protein binds to the SL1 internal loop. For these experiments, we used our SL1m mutant, which is impaired from forming kissing dimers and thus cannot undergo a transition into a duplex dimer (Figure 4-1A). We have shown that replacement of the wild-type GC-rich loop does not impact the integrity of the stem-internal loop-stem element of the SL1 kissing dimer (ref. chapter 2). We performed chemical shift titration experiments to examine NC protein binding to SL1m. 2D-HSQC spectra were recorded with increasing amounts of NC protein up to ratio of SL1m to NC protein of ~1:4 in the presence of 5mM Mg^{2+}. While we did not observe chemical shift perturbations, we observed significant exchange broadening specifically at resonances belonging to the internal loop and neighboring bases (Figure 4-1B and C), including G7, G8, G26, G28 and G31. Thus, these residues either directly interact with NC protein and/or undergo conformational exchange due to NC protein binding. Interestingly, we also observed exchange broadening in the apical loop residue G16, but less so for neighboring G15 and apical loop residue G18. This highlights NC protein's general affinity for consecutive purine residues. Similar results were observed in the absence of Mg^{2+} though broadening effects were somewhat smaller.

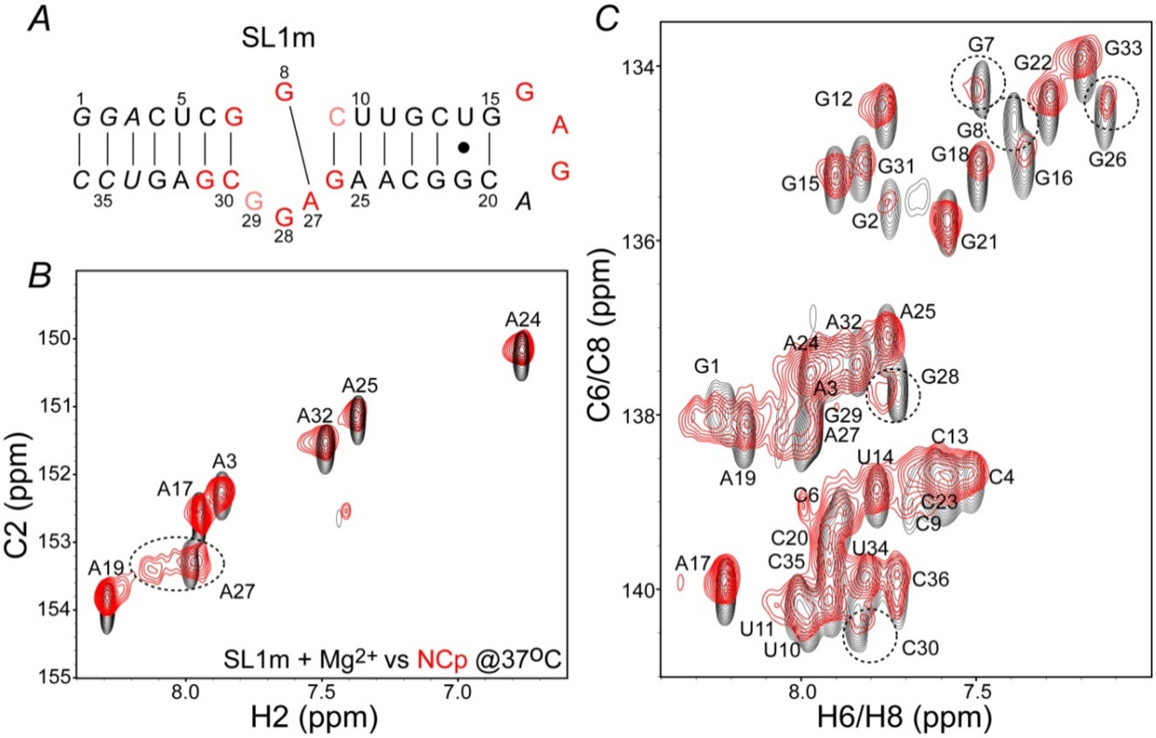

Figure 4-1. Chemical shift mapping of NC protein binding to SL1m in the presence of 5 mM Mg^{2+}. (A) Secondary structure of SL1m (B-C) Overlay of 2D ^{1}H-^{13}C HSQC spectra C2H2 (B) and C6H6 & C8H8 (C) of SL1m recorded in the absence (in black) and presence (in red) of NC protein.

4.2.2 NMR monitoring the NC protein catalyzed isomerization in the presence of Mg^{2+}

As discussed in Chapter 3, spectra of the kissing and duplex SL1 dimers are not identical with several key differences observed for resonances belonging to the apical loop. These resonances afford us the opportunity to study the kissing-to-duplex conversion using time-resolved NMR. None of the residues outside the apical loop show significant differences in chemical shift between kissing and duplex dimers. Thus, at the outset of our study, we had no plans to characterize the dynamic behavior at these sites. For such time-resolved NMR studies, the conversion rate has to be slow compared to the time required to acquire a 2D HSQC spectrum (~22 mins). Especially for SL1, acquiring 2D rather than 1D spectrum was key for resolving resonances. If need be, the rate of conversion can be controlled and optimized by varying three parameters: (i) temperature, (ii) protein:RNA concentration, and (iii) length of the lower SL1 stem. We were able to strike an excellent balance between resolution and sensitivity when working under conditions of ~0.3 mM SL1 kissing dimer, ~0.4 mM NC protein, 8 mM Mg^{2+} at 37°C.

In our experiments, we first recoded 2D HSQC spectra of ~0.3 mM 13C/15N labeled kissing SL1 dimer in the presence of 8 mM Mg^{2+} at 37°C in the absence of NC protein. Next, we performed a series of titrations in which 2D HSQC spectra were recorded following variable incubation with increasing concentrations of unlabeled NC protein. As a control, spectra were also recorded for an identical kissing dimer sample which did not include NC protein. As expected, insignificant changes were observed in the latter control sample throughout the entire time course of our experiments (data not shown).

The addition of NC protein led to broadening of resonances throughout the SL1 kissing dimer consistent with formation of an RNA-protein complex (Figure 4-2). As in SL1m, we observed strong broadening in the internal loop (Figure 4-2 C-F). However, in the kissing dimer, we also observed both significant broadening and perturbations in the apical loop as well as neighboring base-pairs (Figure 4-2C). The latter can arise from both NC protein binding/dissociation and/or any conformational exchange induced by NC protein binding. Interestingly, unlike in SL1m, NC protein induced significant exchange broadening at G21 and to lesser extent at U14 in the U14-G21 base-pair near the apical loop in the kissing dimer (Figures 4-2C). Similar observations have been reported previously by UV melting experiments (1). As described below, the latter broadening as well as some broadening observed in the apical loop was much less significant in the duplex dimer. Thus, NC uniquely interacts with the apical and neighboring residues in the kissing dimer.

Despite the above spectral changes induced by NC protein, we did not observe significant changes in the spectra as a function of time up to RNA:NC-protein concentrations of 2:1.5 at 37°C upon to 12 hours. Specifically, we did not observe a significant increase in the intensity of the characteristic resonances belonging to the duplex dimer apical loop. This indicates that the kissing to duplex dimer conversion is not efficient under these concentrations possibly because more than two NC molecules per kissing dimer are required. Previous studies on kissing dimers containing the internal loop reported that the conversion was highly inefficient at RNA:NC-protein concentrations of 2:2, but was very efficient at ratios of 2:8 (1).

Interestingly, at RNA:NC-protein concentration of 2:2.5, we began observing changes in spectra as a function of time. Denoting this time point as t=0, we recorded 2D HSQC spectra every ~2 hrs for two days and then once every day up to 6 days. In Figure 4-2, we show representative spectra of SL1 recorded at t = ~2 hrs and ~6 days. In these experiments, we observed the gradual disappearance of apical loop resonances characteristic of the kissing dimer and concomitant appearance of corresponding loop resonances characteristic of the duplex dimer. At ~6 days, we did observe a small degree of sample degradation, which is expected given the incubation of the RNA at high temperature (37°C) and presence of divalent ions. The kissing dimer was nearly 100% converted into the duplex dimer after 6 days (Figure 4-2F). The resulting spectra of the SL1 duplex dimer in the presence of NC protein were in very good agreement to spectra obtained when preparing the duplex dimer independently and adding NC protein (Figure

4-2 F and G). These results suggest that a minimum of three NC molecules per kissing dimer are required to induce the kissing to duplex transition. One possibility is that two NC protein molecules are bound to the internal loop and a third to the apical loop.

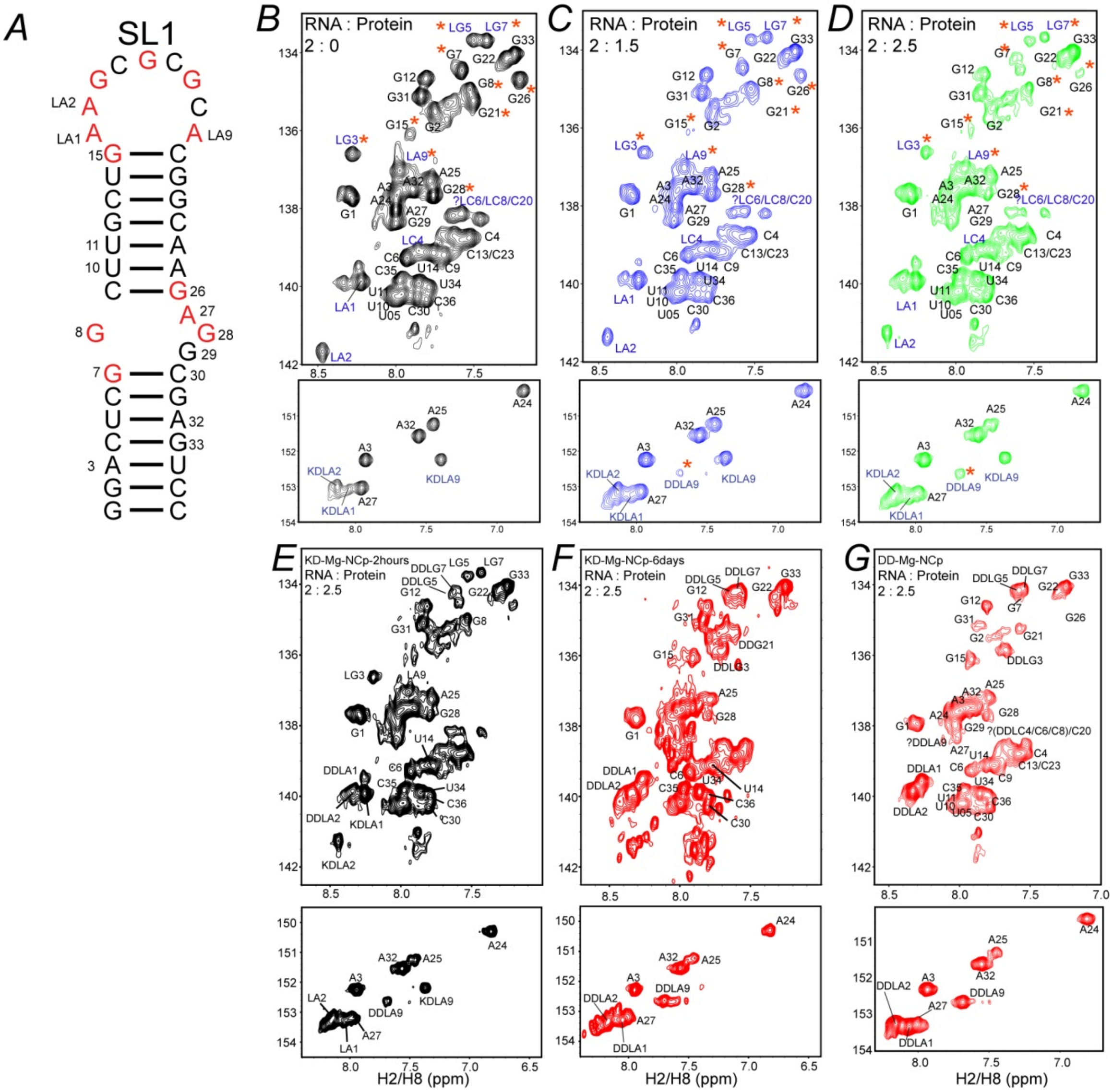

Figure 4-2. SL1 kissing to duplex dimer conversion facilitated by NC protein. (A) Secondary structure of wild-type SL1. Colored residues undergo the greatest degree of exchange broadening upon NC protein binding, (B-D) 2D ^{1}H-^{13}C HSQC spectra of SL1 kissing dimer recorded in the absence of NC protein (B), in the presence of ~0.25mM NC protein (C), in the presence of ~0.4mM NC protein (D), incubation with ~0.4mM NC protein for 2 hours (E), and incubation for 6 days (F). (G) 2D ^{1}H-^{13}C HSQC spectra of SL1 duplex dimer in the presence of NC protein.

4.2.3 Time-resolved NMR analysis of NC protein catalyzed isomerization in the presence of Mg^{2+}: Apical loop resonances

To site-specifically characterize the kinetics of the NC protein-induced kissing to duplex transition, we fitted the time course of individual apical loop resonance intensities to a progress curve (Figure 4-3). Residues from the apical loop can be classified into two types: unpaired purines LA1, LA2 and LA9, which are flexible, and the complementary GCs, including loop G3, G5 and G7, which are essential for the loop-loop interaction. The intensity changes for the resolved apical loop resonances could be fit to a first order exponential curve (Figure 4-3),

$$Y = y_0 + Ae^{-kx} \qquad (1)$$

Strikingly, all apical loop resonances, including those disappearing in the kissing dimer and those appearing in the duplex dimer, yield a similar rate constant of ~5 x 10^{-4} min^{-1} (or ~8 x 10^{-6} sec^{-1}). Thus the data for the apical loop are consistent with a single step conversion between the kissing and duplex dimer.

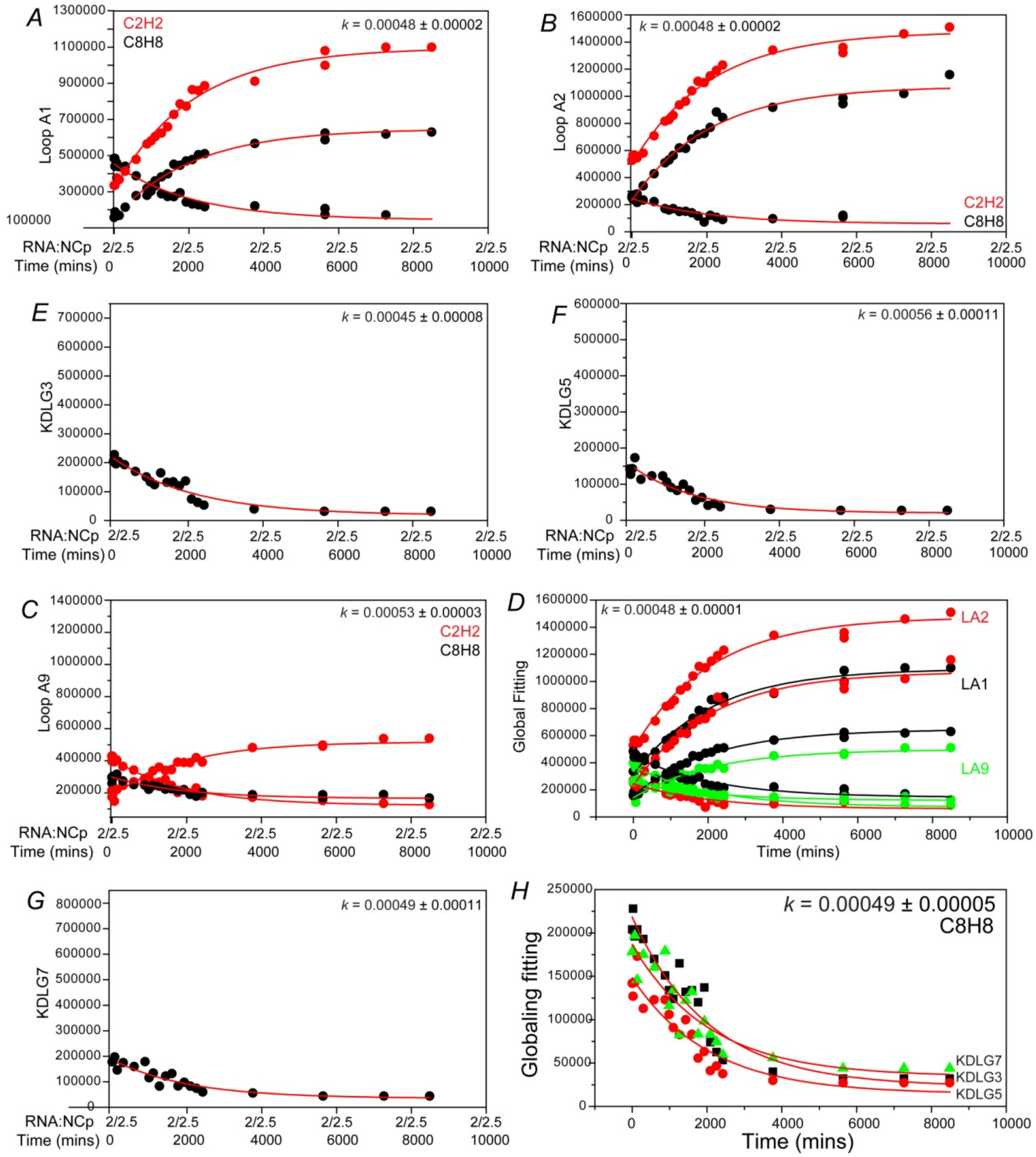

Figure 4-3. The progress fitting curve for the residues in the apical loop (see text for details).

4.2.4 Time-resolved NMR analysis of NC protein catalyzed isomerization in the presence of Mg^{2+}: Non-apical loop resonances

During the time-course of the kissing to duplex conversion, we also observed changes in the intensities of peaks belonging to residues outside the apical loop. In the majority of the cases, the initial broadening induced by NC protein binding prior to t=0 was reduced, which indicates resonance intensities back to an equilibrium value observed for the duplex dime bound to NC protein. These changes in resonance intensities must be interpreted with great caution since they do not simply represent the population of a given species. Rather, they probed a mixture of both kissing and duplex populations at t=0 as well as on exchange processes involving NC binding/dissociation and conformational changes. Nevertheless, we examined these resonances individually and characterized their progress curves.

With the exception of resonances belonging to the U14-G21 base-pair near the apical loop, we did not observe significant changes in the resonance intensities of five resolved resonances belonging to four residues in the upper stem. For U14, we observed a diminishment in the exchange broadening induced by NC protein binding. Interestingly, the increase in resonance intensity followed a first order exponential curve yielding a rate constant ~1.8 x 10^{-4} min^{-1} which is less than half of that observed for the apical loop resonances (Figure 4-4). The data from U14 could not satisfactorily be combined with the data from apical loop in a global fitting. Notwithstanding the complications mentioned above, the slower rate observed for U14 might be expected given that unlike the apical loop, the intra-monomer U14-G21 hydrogen bond in the kissing dimer must be disrupted in order to form the corresponding inter-monomer hydrogen bond in the duplex dimer. In contrast, the G21 resonance from the kissing dimer vanished during the NC protein titration, and the corresponding duplex G21 resonance appeared subsequently in a crowded region of the spectrum (Figure 4-2).

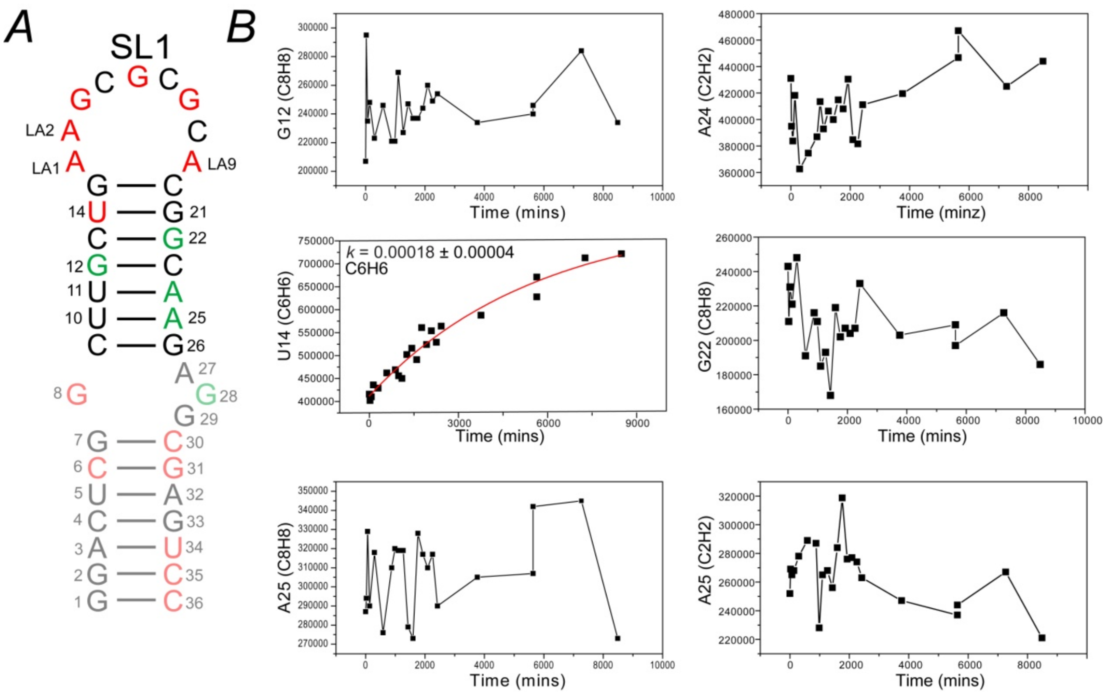

Figure 4-4. The intensity fluctuations of the upper stem residues during the kissing to duplex conversion.

We now turn our attention to the internal loop. For A27, we observe small complex changes over time (Figure 4-5). In contrast, we observed a significant decrease in the resonance intensity for both G28 and G8 which form a mis-match with A27 in SL1 monomer (2). The variations for G28 were complex (Figure 4-5). In contrast, the changes for G8 fit reasonably well to a first order exponential curve yielding a rate constant ~5 x 10^{-4} min^{-1} similar to that observed for the apical loop (Figure 4-5). Unfortunately, the C8 resonance of A27 and G29 are on top of each other, so we were not able to follow their changes over time.

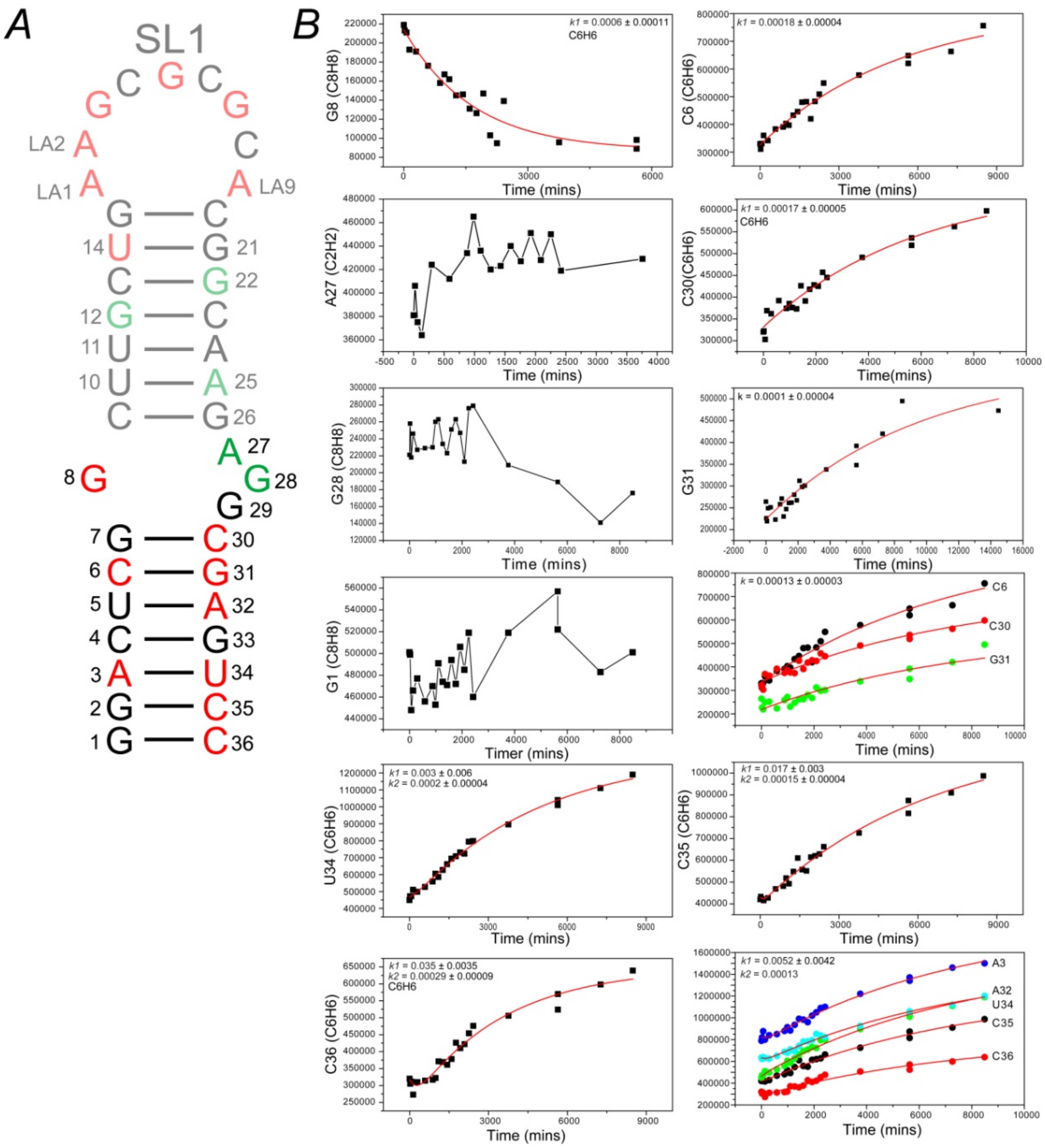

Figure 4-5. The intensity tracking of the lower stem of the kissing dimer.

Proceeding down to the lower stem, we observed insignificant changes in the intensity of terminal residue G1. In contrast, we observed a marked increase in the resonance intensities for a number of residues in the lower stem. Residues C6, C30, and G31 could also be fit to a first order exponential curve yielding a rate constant ~1-2 x 10^{-4} min^{-1} which is similar to values observed in other Watson-Crick base-paired residues within helical stems but smaller than values observed for the internal and apical loop. The data from residues C6, G30, and C31 could be globally fit to obtain a single rate constant of 1.3 x 10^{-4} min^{-1}.

Further down the helix, residues A3, A32, C35, and C36 appeared to exhibit a lag phase. To verify the existence of the lag phase, we fitted the data of the lower stem residues using both first order (1) and second order exponential equations (2) as shown below,

$$Y = y_0 + A_1 e^{-k1x} + A_2 e^{-k2x} \tag{2}$$

In Figure 4-6, we show the difference between the measured and back-predicted intensities when using first (Figure 4-6A) and second (Figure 4-6C) order exponential equations. The measured intensities in the early phase are systematically overestimated when using the first order equation whereas this is not the case when using the second order equation (Figure 4-6 B and D). While the second order exponential fittings agree best with the data, a good fit is also obtained with a single first order exponential fit.

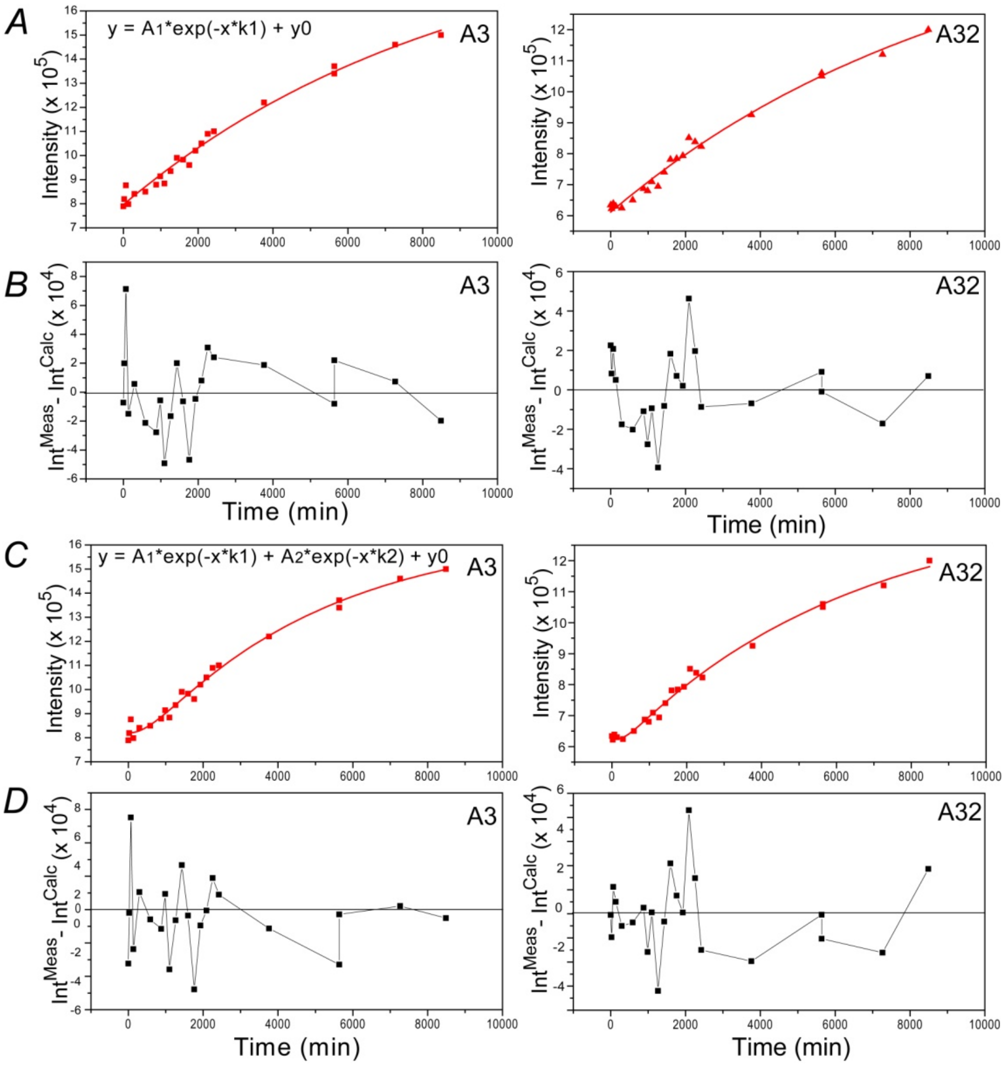

Figure 4-6. Characterization of lag phase of individual residues A3 and A32. (A) the first order exponential fitting, (C) the second order exponential fitting, (B and D) the differences of fitted value and original value.

The “slow” kinetic rate constant obtained for A3, A32, C35, and C36 ranges is close to ~2 x 10^{-4} min^{-1} and is similar to rate constants observed for other residues in the lower stem. In fact, a good fit was observed to the data of A3, A32, C35, and C36 when setting one of the rate constants equal to that obtained from fitting the data from residues C6, C30, and G31. The “fast” rate constant ranges around ~3 x 10^{-3} min^{-1} which is roughly one order of magnitude faster than the value observed for the “slow” rate.

4.3 Discussion

Our study provided new insights into the kissing to duplex transition and the interaction between NC protein and SL1. The chemical shift mapping experiments using both SL1m and SL1 kissing and duplex dimers clearly demonstrate that NC protein binds to the SL1 internal loop inducing exchange broadening. The intermediate exchange broadening indicates relatively weak binding ($K_d \sim \mu M$) consistent with previous SL1 NC protein binding studies (2-5). Binding of NC protein to the SL1 internal loop is also in agreement with a mass spectrometry study which showed not only that NC protein binds the internal loop but also that internal loop bound NC proteins are the ones which activate the kissing to duplex transition (6). Interestingly, the internal loop G26/A27/G28/G29 sequence is identical to the apical loop sequence of SL3 which had been shown to have the tightest binding to NC protein (3,7).

Chemical shift titrations involving the kissing SL1 dimer revealed additional potential interactions between NC protein and SL1. As in SL1m, significant broadening was observed at the internal loop. However, we also observed significant exchange broadening at several residues in the apical loop (Figure 4-2). The broadening at the apical loop likely reflects NC protein binding though we cannot rule out the added possibility that the broadening arises from conformational exchange. Importantly, though broadening of the internal loop is observed in duplex SL1, we do not observe significant broadening in corresponding apical loop resonances. Furthermore, the addition of NC protein led to broadening at the U14-G21 base-pair in the upper stem of kissing SL1 but not in either SL1m or the duplex dimer. A recent 1D-NMR study also noted that the uridine imino proton in the U14-G21 base-pair in the kissing dimer was significantly broadened by the presence of NC protein (1). It is possible that the labile kissing apical loop renders this base-pair flexible creating a unique NC protein binding site comprising the apical loop and neighboring residues. This could also explain the requirement for a minimal NC protein to RNA stoichiometry of 3:2 for efficient kissing to duplex conversion. The SL1-NC protein complex formed prior to conversion into the duplex may have alternative structural features, and perhaps even correspond to a G-quadruplex intermediate that was proposed in our and other studies (8).

Following formation of the NC protein-SL1 complex, the apical loop shows a very clear one step kissing to duplex conversion with a rate constant of ~5 x 10^{-4} min^{-1}. Similar rate constants were observed for internal loop residue G8. The small rate constants observed are consistent with a previous study using the same SL1 construct in which the kissing to duplex conversion did not take place after 18 hours at an RNA:NC protein ratio of 2:2 at ambient temperature. In the HIV particle, the much higher NC protein:RNA ratio (7nt of RNA per NC protein) will likely ensure that the conversion occurs significantly faster and on a biologically relevant timescales. In our study, we had to slow down the reaction to allow time-resolved studies by NMR and this is clearly one of the weaknesses of our approach.

Lower (2-fold) first order kinetic rate constants were observed for Watson-Crick base-pairs outside the internal and apical loop. The quantitative interpretation of these data is however less straightforward since the intensities do not simply reflect the population of species at a given time point but rather can also be affected by exchange broadening effects. Notwithstanding the latter, the slower rates observed may indicate that regions outside the apical and internal loop undergo the transition semiindependently with their own characteristic rate constants. The slower rates are expected given that for helical residues, hydrogen bonds have to be disrupted within a given monomer prior to strand exchange and structural isomerization. The similar rate constants observed for these helical residues suggests that their transition may occur through a common single step. Though we observe possible evidence for a lag phase in a number of residues in the lower stem, consistent with an additional rapid process, the interpretation of these data remains complicated by the exchange broadening effects described above.

Taken together, our data along with previous published results are consistent with a model in which SL1 is in rapid equilibrium with an NC protein bound state which likely involves three NC protein molecules bound to the internal and apical loop. Our data do not provide direct information regarding the conformation of this intermediate and it could involve an altered SL1 conformation, such as the quadruplex geometry proposed previously (9) and visualized by molecular simulations in the context of a short SL1 kissing dimer, which contains the apical loop and upper stem (10).

In this conformation, all four strands are in alignment for strand exchange (Figure 4-7). This NC protein bound conformation is capable of undergoing a transition towards the duplex dimer and proceed efficiently once the NC-protein:RNA concentration is >2.

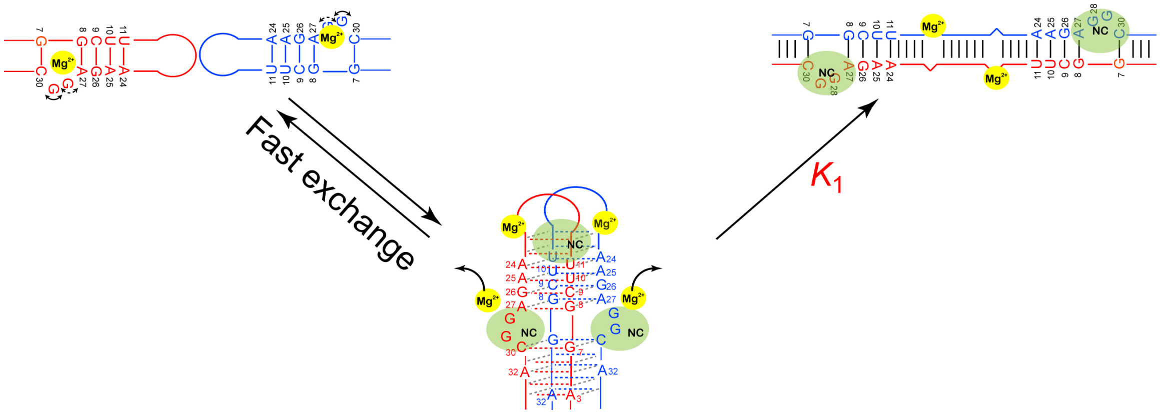

Figure 4-7. Proposed role for bulge-induced dynamics and NC protein-dependent SL1 structural transition between the kissing and duplex dimers.

4.4 Preparation, purification, and characterization of nucleocapsid protein and RNA samples for NMR studies

4.4.1 RNA synthesis

Uniformly $^{13}C/^{15}N$ labeled SL1 and individually labeled SL1 (GC-SL1 and AU-SL1) NMR samples (0.6-1.0mM) were prepared using standard in vitro transcription reactions using T7 RNA polymerase from synthetic DNA templates. SL1 was purified by 20% PAGE and electroelution by Elutrap. The sample was buffer exchanged, concentrated by centricon, and stored at -20°C. The buffer used in NMR experiments contained 15 mM sodium phosphate, 25 mM sodium chloride, 0.1 mM EDTA, and pH ~ 6.4 in H2O. RNA concentration was quantified by ultraviolet (UV) spectrophotometry and stored at -20°C.

4.4.2 Kissing and duplex dimer preparation and assay

In the case of the kissing dimer sample preparation, a 1~5μM RNA solution in $H_2O$was heated to 90°C for 10 mins and immediately cooled in ice-water for 30mins.

Then RNA sample was exchanged to NMR buffer and concentrated to ~1mM. For duplex preparation, a 1mM RNA sample in NMR buffer was heated to 90°C for 10 mins and incubated at room temperature for 1 hour. Both RNA dimer samples were checked by running 10% native PAGE in TBE or TBM buffer at 4° C. The bands of kissing dimer ran at the same position as duplex dimer in TBM as the dimeric formation; however in TBE it ran as a monomer, migrating much faster than the duplex dimer (1,11,12).

4.4.3 NC protein expression

The plasmid we used for NC protein expression was a generous gift from Dr. Gaya Amarasinghe (3,4,13). The plasmid was constructed by sub-cloning the NC protein coding sequence from the infectious HIV-1_{NL4-3} clone into the expression vector pET-3a (Novagen) using a polymerase chain reaction based strategy. To amplify and ensure the absence of errors, the plasmid was transformed into competent Nova Blue Escherichia coli and ampicillin resistant colonies were selected. The plasmid was isolated by a Qingan plasmid purification kit and was sequenced by using T7 promoters as primers.

For protein expression, the pET-3a-NC protein was transferred to BL21(DE3)LysS E.coli. and grown for overnight at 37°C on an LB medium agar plate in the presence of ampicillin (100ug/mL) and chloramphenicol (34ug/mL). Single transformed colonies were inoculated into 50ml starting medium and grown for 10 to 12 hours at 30 °C. The culture was then diluted 10 fold and grown at 37 °C to an OD_{600} = 0.6. Protein expression was induced by IPTG (isopropy1-thio-β,D-galactopyranoside) with a final concentration of 1mM. Cells were harvested by centrifugation after an additional 4 hours of growth at 37 °C, and cell paste was stored at -80 °C.

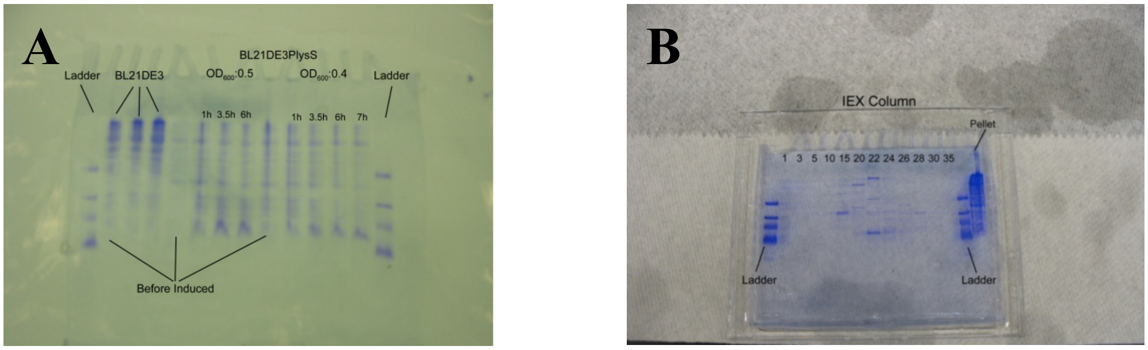

Figure 4-8. PAGE gel analysis of the yield and purity of NC protein. (A) Optimization of NC protein expression and (B) cation exchange column to purify NC protein.

To prevent NC protein oxidation and degradation, all of the following steps of purification were conducted at 4°C and all buffers were degassed and contained $ZnCl_2$ and β-mercaptoethnal. A ~5 gram cell paste from 5L of culture was suspended in 30ml lysis buffer (75µM $ZnCl_2$, 50mM Tris-HCl (pH: 9.5), 50mM NaCl, 10% glycerol) with proteinase inhibitors (35µg/ml phenylmethyl sulfonyl fluoride, 0.3mg/ml EDTA, 0.5µg/ml leupeptin A, and 0.7µg/ml pepstation A). Cell lysis was achieved by passing suspended cell paste through a French press twice. The lysate was mixed with 0.3%-0.4% polyethylenimine on ice for 10 mins and was centrifuged twice. The clear supernatant was loaded onto 25mL SP-sepharose column with a flow rate of ~1ml/min. Elution of NC protein was achieved with a 100ml linear gradient of 25mM-2M NaCl in lysis buffer. Fractions containing NC protein were detected by Tris-Tricine SDS-polyacrylamid gel electrophoresis (PAGE) (Figure 4-11). Fractions were pooled and concentrated using a centricon (3KDa cutoff) to 1ml. The protein solution was then purified by a G50 size exclusion column with a 0.5ml/min flow rate by using GF buffer (25mM NaCl, 10% glycerol, 15mM sodium phosphate pH 6.8, 75µM $ZnCl_2$). The peak fractions were pooled and concentrated by centricon again. The integrity and purity of NC protein were checked by SDS-PAGE and mass spectrometry. The final concentration of protein was determined using the Bradford protein assay (Bio-Rad) with bovine serum albumin as the standard.

4.4.4 Mass spectroscopy assay

The experiments were conducted on an actively shielded 7T Fourier Transform Ion Cyclotron Resonance (FT-ICR) mass spectrometer (Bruker Daltonics, Billerica, MA) in positive ion mode. NC protein was dissolved in 20% Methanol with 0.1% formic acid to a concentration of 1 μM and infused via an external Apollo II ion funnel electrospray ion source at a flow rate of 50 μL/h with the assistance of N_2 nebulizing gas to generate multiply charged precursor cations. Following ESI, ions were externally accumulated for 1s in a hexapole, transferred through high-voltage ion optics, and captured by gated trapping in an infinity ICR cell. This accumulation/transfer/trapping sequence was looped twice for an improved signal-to-noise ratio. All mass spectra were acquired with XMASS (version 6.1, Bruker Daltonics) in broadband mode from m/z 200 to 2,500 with 512k data points and data processing was performed with the MIDAS analysis software (14). Several charge states were observed for NC protein and the 10+ and 9+ charge states species were used for calculating the mass (results shown in the following table).

Table 4-1. Calculated mass for NC protein using 10+ and 9+ charge state species. Average mass is calculated by considering the relative abundance of all the isotopic peaks of the entire isotopic distribution envelope.

NC protein	Monoisotopic mass (Da)	Average mass (Da)
10+	6359.072	6364.122
9+	6359.075	6364.397

4.4.5 NMR Spectroscopy

All NMR experiments were preformed on an Avance Bruker 600 MHz NMR spectrometer equipped with a triple-resonance 5 mm cryogenic probe. Unless stated otherwise, all NMR experiments were conducted at 298 K. NMR spectra were processed using NMRPipe/NMRDraw, analyzed using NMRView, and overlaid using SPARKY 3. All non-exchangeable resonances were assigned using standard methods and found to be in excellent agreement with those reported previously for the kissing and duplex construct. We assigned the non-exchangeable resonances from the complex of NC protein with SL1 kissing and with duplex dimers based on our SL1 assignments (ref chapter 3).

Experiments for tracking the kissing to duplex conversion were performed by recording 2D 1H-13C HSQC spectra continuously following gradually increases the NC protein to RNA ratio by titrating NC protein into RNA sample (RNA:protein = 2:0.1, 2:0.5, 2:1, 2:1.5, 2:2.5) at 37 ^{o}C; subsequently, we constantly recorded 2D 1H-13C HSQC spectra at fixed RNA to NC protein ratio (2:2.5) until kissing dimer converted to duple conformation (Figure 4-9).

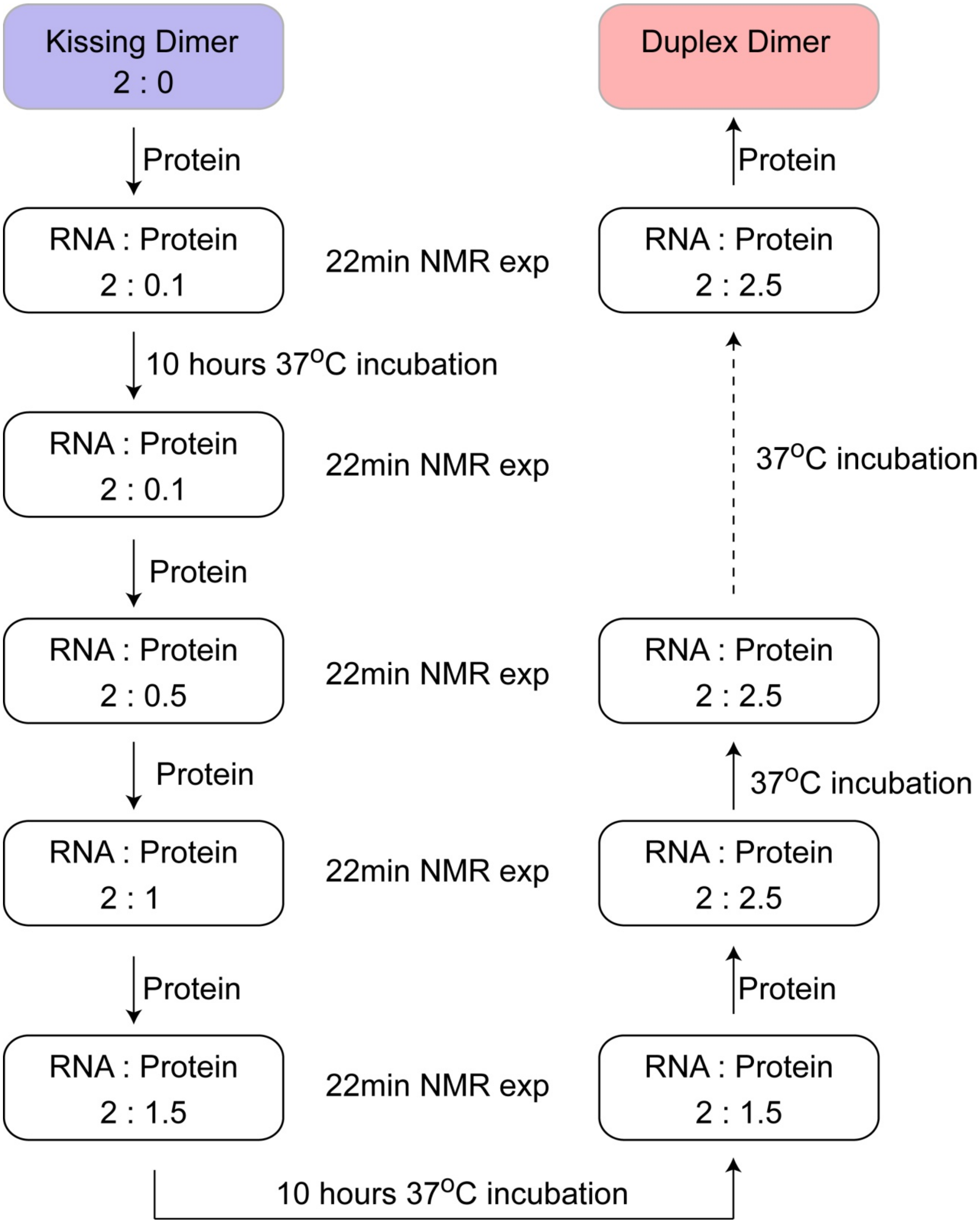

Figure 4-9. Experiment design for tracking the kissing to duplex transition in the presence of NC protein by NMR spectroscopy.

4.5 References

1. Mujeeb, A., Ulyanov, N.B., Georgantis, S., Smirnov, I., Chung, J., Parslow, T.G. and James, T.L. (2007) Nucleocapsid protein-mediated maturation of dimer initiation complex of full-length SL1 stemloop of HIV-1: sequence effects and mechanism of RNA refolding. *Nucleic Acids Res*, **35**, 2026-2034.
2. Lawrence, D.C., Stover, C.C., Noznitsky, J., Wu, Z. and Summers, M.F. (2003) Structure of the intact stem and bulge of HIV-1 Psi-RNA stem-loop SL1. *J Mol Biol*, **326**, 529-542.
3. De Guzman, R.N., Wu, Z.R., Stalling, C.C., Pappalardo, L., Borer, P.N. and Summers, M.F. (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science*, **279**, 384 - 388.
4. Amarasinghe, G.K., De Guzman, R.N., Turner, R.B., Chancellor, K.J., Wu, Z.R. and Summers, M.F. (2000) NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *J Mol Biol*, **301**, 491-511.
5. Shubsda, M.F., Paoletti, A.C., Hudson, B.S. and Borer, P.N. (2002) Affinities of packaging domain loops in HIV-1 RNA for the nucleocapsid protein. *Biochemistry*, **41**, 5276-5282.
6. Hagan, N.A. and Fabris, D. (2007) Dissecting the protein-RNA and RNA-RNA interactions in the nucleocapsid-mediated dimerization and isomerization of HIV-1 stemloop 1. *J Mol Biol*, **365**, 396-410.
7. D'Souza, V. and Summers, M.F. (2005) How retroviruses select their genomes. *Nat Rev Microbiol*, **3**, 643-655.
8. Awang, G. and Sen, D. (1993) Mode of dimerization of HIV-1 genomic RNA. *Biochemistry*, **32**, 11453-11457.
9. Rist, M.J. and Marino, J.P. (2002) Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1. *Biochemistry*, **41**, 14762-14770.
10. Aci, S., Mazier, S. and Genest, D. (2005) Conformational pathway for the kissing complex-->extended dimer transition of the SL1 stem-loop from genomic HIV-1 RNA as monitored by targeted molecular dynamics techniques. *J Mol Biol*, **351**, 520-530.
11. Baba, S., Takahashi, K., Nomura, Y., Noguchi, S., Koyanagi, Y., Yamamoto, N., Takaku, H. and Kawai, G. (2001) Conformational change of dimerization initiation site of HIV-1 genomic RNA by NCp7 or heat treatment. *Nucleic Acids Res Suppl*, 155-156.
12. Takahashi, K.I., Baba, S., Chattopadhyay, P., Koyanagi, Y., Yamamoto, N., Takaku, H. and Kawai, G. (2000) Structural requirement for the two-step dimerization of human immunodeficiency virus type 1 genome. *Rna*, **6**, 96-102.
13. Amarasinghe, G.K., Zhou, J., Miskimon, M., Chancellor, K.J., McDonald, J.A., Matthews, A.G., Miller, R.R., Rouse, M.D. and Summers, M.F. (2001) Stem-loop SL4 of the HIV-1 psi RNA packaging signal exhibits weak affinity for the nucleocapsid protein. structural studies and implications for genome recognition. *J Mol Biol*, **314**, 961-970.

14. Mo, J. and Hakansson, K. (2007) Oligonucleotide gas-phase hydrogen/deuterium exchange with D2S in the collision cell of a quadrupole-Fourier transform ion cyclotron resonance mass spectrometer. *Anal Chem*, **79**, 7893-7898.

Chapter 5

Conclusion and future directions

5.1 Conclusion

The dimerization and structural isomerization of SL1 is a hierarchical dynamical process that involves the complex interplay of intra-molecular and inter-molecular structural dynamics, Mg^{2+} association, and NC protein binding. In order to dissect different aspects of SL1 that are key for dimerization and isomerization, we first examined the conformation of individual monomeric units by using a simplified SL1 mutant that is impaired from forming dimers by replacement of the wild-type palindromic loop with a GAGA tetraloop. In working with this mutant, our primary focus was the highly conserved but poorly characterized AGG-G internal loop that is critical for NC protein binding and isomerization. We quantitatively characterized internal motions induced by this internal loop and its Mg^{2+} binding properties using a variety of state-of-the-art modern NMR techniques including residual dipolar couplings (RDCs) for probing motions at sub-millisecond timescales, domain-elongation and spin relaxation measurements to probe motions at picosecond-to-nanosecond timescales, and chemical shift mapping to characterize Mg^{2+} binding.

By monitoring temperature-dependent exchange broadening of resonances and J_{NN}-COSY cross-peaks, we obtained direct evidence for a previously hypothesized inter-conversion between alternative SL1 secondary structures, which causes the partial melting and re-annealing of three base pairs in the SL1 upper stem and internal loop. Our domain elongation studies revealed diffusion-limited inter-domain motions occurring about the internal loop at nanosecond timescales. We showed for the first time that Mg^{2+} binds specifically to the internal loop region and arrests the secondary structural conversion, thus stabilizing base pairs in the upper stem. While Mg^{2+} binding quenches inter-domain motions, it apparently leads to an increase in the local dynamics of the internal loop residues G28 and G29 which are believed to bind NC protein. We proposed that Mg^{2+} could serve to ensure that the kissing-duplex transition only occurs in the presence of NC protein, thus ensuring that viral RNA and protein maturation occur in concert. Furthermore, we proposed a model for the kissing-to-duplex conversion involving formation of a quadruplex geometry in which the internal loop acts as a key nucleation site for strand exchange.

With the characterization of the SL1 monomer, we turned our attention to the larger, more complex, and spectroscopically more challenging kissing and duplex dimers. By monitoring the chemical shift perturbation and measuring exchange broadening at different temperatures, we showed that the dynamical and Mg^{2+} binding properties of the monomer SL1 are retained in the kissing dimer. However, we observed differences between the dynamical and Mg^{2+} binding properties of the apical loop, with the kissing dimer generally being more flexible. In particular, in the kissing dimer, the 5' unpaired adenines undergo a slow μs-ms process likely corresponding to a looping-in-and-out motion which is dampened by Mg^{2+} binding. In contrast, the same adenines undergo faster ps-ns local motions in the duplex dimer likely involving a looped out conformation and are not significantly affected Mg^{2+} binding. The mobility at the adenines in the kissing apical loop likely affords a hinge that permits individual monomeric units to come into close proximity without disrupting the loop-loop interaction and also play a role in NC binding.

Finally, we used time-resolved NMR methods to characterize the NC protein-dependent isomerization between kissing and duplex dimers. We investigated NC protein binding to the kissing and duplex dimer in the absence and presence of Mg^{2+}. Using site-specific chemical shift mapping experiments, we were able to identify a primary NC protein binding site in the internal loop of the kissing dimer. We obtained evidence that NC stabilizes an intermediate conformation prior to converting the kissing dimer into a duplex dimer through what seems to be a single step conversion. Based on our earlier characterization of melting dynamics induced by the bulge, we proposed that the above intermediate consists of two SL1 monomers with their four strands in a quadruplex geometry in which base-pairs above the internal loop in a given monomer are breaking as they form inter-molecular base-pair alignments. In this model, Mg^{2+} stabilizes SL1 conformation but also enhances NC protein binding by looping out key internal loop residues, thus increasing the dependency of isomerization on NC protein binding. Together, our results revealed that a complex interplay of intrinsic SL1 dynamics, Mg^{2+}, and NC protein binding allow the SL1 kissing dimer to convert into a duplex dimer in a highly selective NC protein-dependent pattern. This may help ensure that viral RNA and protein maturation occur in a highly concerted fashion.

5.2 Future direction

An important future goal is to compare the structural dynamics as well as Mg^{2+} , protein binding properties of SL1 and the kissing to duplex transition in the context of the entire packaging signal and the HIV-1 5’ untranslated region. Methods are available to tackle such large RNAs by NMR using a combination of isotope-edited approaches (1,2) and segmental labeling of SL1 into an otherwise unlabeled HIV-1 5’ RNA leader(3,4). These studies are necessary to explore dimerization and isomerization within the native HIV-1 leader context in which long-range tertiary interactions could play an important role. It would also be of great interest to repeat the time-resolved kinetic studies of isomerization from the perspective of the NC protein. This could be carried out by repeating the studies using isotopically labeled NC protein and unlabeled SL1 and fast NMR methods for data acquisition for instance, SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds (5). These studies could provide insights into the number of NC proteins that associate with SL1 as well as elucidate conformational changes that might take place during the isomerization reaction coordinate.

As mentioned in chapter one, besides initiating the HIV-1 gRNA dimerization, SL1 is believed to be involved in another large RNA conformational switch between a branched (BMH) and linear (LDI) conformation, which is also catalyzed by NC protein. This conformation switch could play a role in temporally regulating different viral functions. Existing evidence for these alternative conformations comes from qualitative biochemical analysis of gel mobility (6-8). A recent study on the RNA-DNA chimera reported that NC protein recognizes and binds a GUC sequence (9). Interestingly, in the two alternative HIV-1 conformations, a short GUG bulge exists in the BMH conformation but is masked in a helix conformation in LDI, which surprisingly interacts with the key GC rich complementary apical loop of SL1 (figure 5-1). Such a conformational switch may play an important role in regulating intracellular gene expression functions or virion association function. Moreover distinct conformations of HIV-1 5’ leader RNA may necessitate the dual role of HIV-1 transcripts as message RNA and viral genome (10,11). Future studies could provide new insights into these conformational transitions that involve large fragments of the HIV-1 genome and ultimately bridge the scales between what we see on the individual domain level and the entire multi-domain assembly.

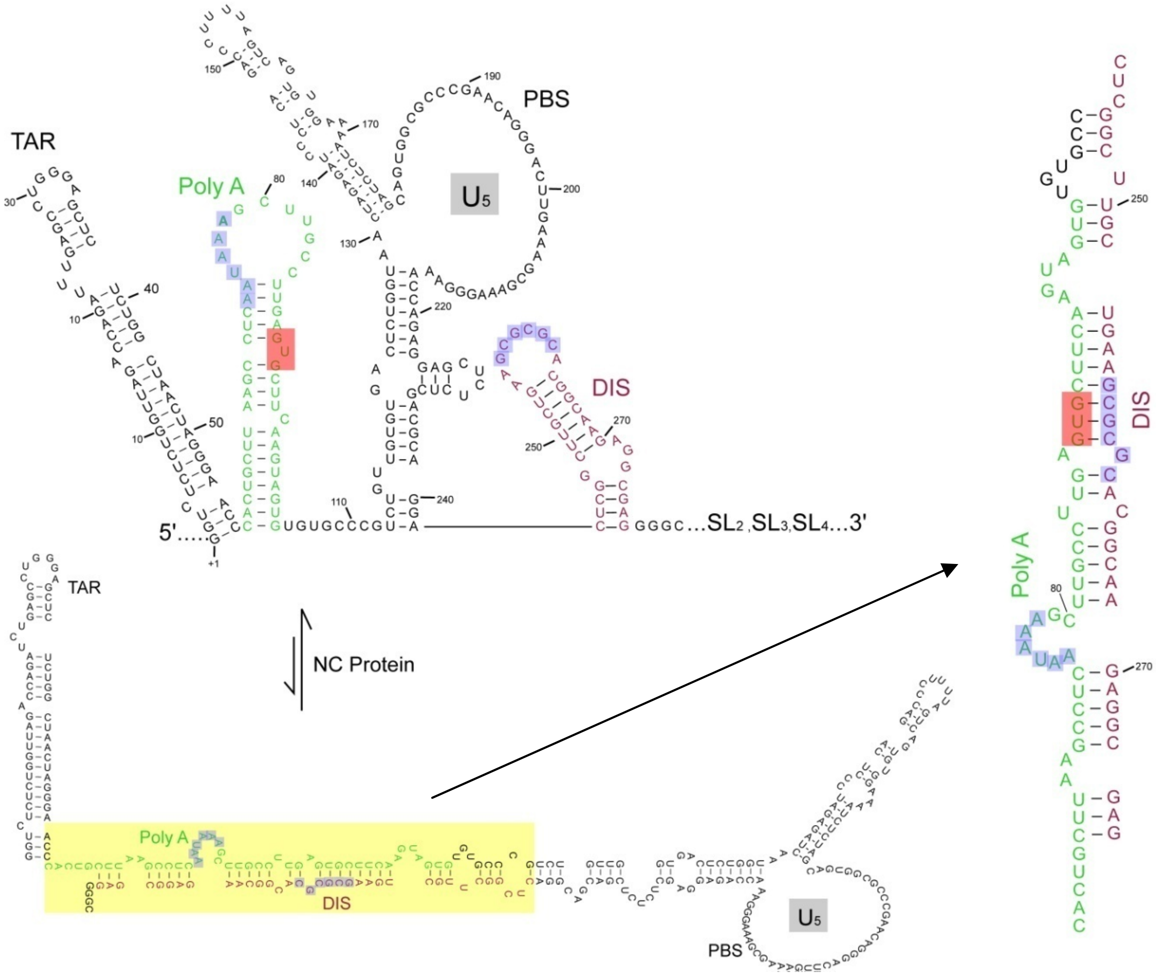

Figure 5-1. Two alternative conformations of 5' UTR of HIV-1

5.3 References

1. D'Souza, V. and Summers, M.F. (2005) How retroviruses select their genomes. *Nat Rev Microbiol*, **3**, 643-655.
2. D'Souza, V., Dey, A., Habib, D. and Summers, M.F. (2004) NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus. *J Mol Biol*, **337**, 427-442.
3. Kim, I., Lukavsky, P.J. and Puglisi, J.D. (2002) NMR study of 100 kDa HCV IRES RNA using segmental isotope labeling. *J Am Chem Soc*, **124**, 9338-9339.
4. Kim, N.K., Murali, A. and DeRose, V.J. (2004) A distance ruler for RNA using EPR and site-directed spin labeling. *Chem Biol*, **11**, 939-948.
5. Schanda, P., Kupce, E. and Brutscher, B. (2005) SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR*, **33**, 199-211.
6. Ooms, M., Huthoff, H., Russell, R., Liang, C. and Berkhout, B. (2004) A riboswitch regulates RNA dimerization and packaging in human immunodeficiency virus type 1 virions. *J Virol*, **78**, 10814-10819.
7. Abbink, T.E., Ooms, M., Haasnoot, P.C. and Berkhout, B. (2005) The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. *Biochemistry*, **44**, 9058-9066.
8. Das, A.T., Vink, M. and Berkhout, B. (2005) Alternative tRNA priming of human immunodeficiency virus type 1 reverse transcription explains sequence variation in the primer-binding site that has been attributed to APOBEC3G activity. *J Virol*, **79**, 3179-3181.
9. DeCiantis, C.L., Jensen, D.K., Hudson, B.S. and Borer, P.N. (2007) A Nucleic Acid Switch Triggered by the HIV-1 Nucleocapsid Protein. *Biochemistry*, **46**, 9164-9173.
10. Huthoff, H. and Berkhout, B. (2001) Two alternating structures of the HIV-1 leader RNA. *Rna*, **7**, 143-157.
11. Huthoff, H. and Berkhout, B. (2001) Mutations in the TAR hairpin affect the equilibrium between alternative conformations of the HIV-1 leader RNA. *Nucl. Acids Res.*, 29, 2594-2600.